UNIVERSITE DE BLIDA 1

Faculté de Technologie

Département d'Electronique

THESE DE DOCTORAT EN SCIENCES

en Electronique

DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS

Par

KARA MOSTEFA KHELIL Chérifa

Devant le jury composé de :

A. FERDJOUNI	MC. A, U. de Blida 1	Président
L. HASSAINE	DR, CDER, Alger	Examinatrice
Y. BOT	MC. A, U. de Khemis Miliana	Examinateur
K. KARA	Professor, U. de Blida 1	Rapporteur
A. CHOUDER	Professor, U. de M'sila	Co-Rapporteur
B. AMROUCHE	MC. A, U. de Blida 1	Invitée

Blida le 17/07/2021

DEDICATION

This dissertation is wholeheartedly dedicated to my beloved parents, my father ***Majid***, my mother ***Hadjira*** who have been my source of inspiration and gave me strength when we thought of giving up, who continually provide their moral, spiritual and emotional Support.

To my brothers ***Bachir***, ***Hachemi***, ***Islem*** and ***Sid Ali***, who have never left my side and are very special.

I also dedicate this dissertation to my husband, ***Mohamed TEFFAHI***, who has been a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life.

To my wonderful children ***Mohamed Mohcen*** and ***Youcef Wassim***, as well as, ***Nour Marwa Aya*** and ***Malak Raounak*** who have been affected in every way possible by this quest.

I am grateful that you are all kept me going and this work not have been possible without you.

Thank you very much god bless you

ACKNOWLEDGEMENTS

First and foremost, I would like to express gratitude to ***Allah SWT***, for all opportunities, blessings, trials, inspirations and strengths that have been showered on me to accomplish this thesis.

Second, from the bottom of my heart I would particularly like to say big thanks to Doctor ***Badia AMROUCHE*** who guided me so positively, who made me feel confident in my abilities after a long moment of despair and who was always willing and enthusiastic to assist.

Furthermore, I would like to thank my supervisors, Professor ***Kamel KARA*** and Professor ***Aissa CHOUDER*** for their consideration guidance as well as their consistent support during the running of this research.

I would also like to acknowledge Doctor ***Abou Soufiane BENYOUCEF*** for his participation and engagement in this study.

I am extremely grateful to my reviewers: Doctor ***Abdelaziz FERDJOUNI***, Doctor ***Linda HASSAINE*** as well as Doctor ***Youcef BOT***, for taking part of their time to read this thesis as well as for their valuable comments and suggestions.

To conclude, I cannot forget the heartfelt thanks to my family and friends for all the unconditional support in these very intense academic years.

Kara Mostefa Khelil Cherifa

2021

ABSTRACT

In this thesis, a new diagnosis method for Photovoltaic (PV) systems using Artificial Intelligence is developed. This diagnosis method uses the Back Propagation Artificial Neural Networks (BPNN) classifiers to detect and diagnose faults in the Direct Current (DC) side of Grid Connected Photovoltaic (GCPV) systems. For this, two BPNNs are required: the first ANN is necessary for the classification of current from two input data (solar irradiation level and current at maximal power point), while the second ANN is called for voltage classification from the two input data (cell temperature and voltage at maximal power point). The output of both ANNs represent the input of a combinational algorithm in order to obtain the diagnosis of the PV generator characterizing the final step of this approach. This algorithm diagnoses the most frequent faults encountered in PV installations that are: one short-circuited module in PV generator, two short circuited modules in PV generator, four short circuited modules in PV generators and one disconnected string in PV generator. The obtained results of the cited method are excellent with an average overall accuracy of 98.6%.

In the reason to find the best choice of ANNs, five types of algorithms have been tested that are: Back Propagation Neural Network (BPNN), Probabilistic Neural Network (PNN), Generalized Regression Neural Network (GRNN) and two Radial Basis Function Neural Network (RBF). These ANNs have been tested and compared using the same faults and the same work conditions. The obtained results brought a good clarification and demonstrate that the PNN algorithm takes the top of the list as the best ANN from the point of view of its response time as well as its displaying of 100% in all statistical concepts comparing to other algorithms. The efficiency of the developed method is experimentally evaluated by using real measured data, collected from real GCPV system located at the Centre Des Energies Renouvelables in Algiers (Algeria).

Keywords: Grid connected PV plant, faults detection, diagnosis, Artificial Neural Network (ANN), working conditions, maximum power point (MPPT).

RESUME

Dans cette thèse, une nouvelle méthode de diagnostic des systèmes photovoltaïques (PV) utilisant l'intelligence artificielle est développée. Cette méthode de diagnostic utilise les classificateurs BPNN (Back Propagation Neural Network) pour détecter et diagnostiquer les défauts du courant continu (DC) des systèmes photovoltaïques connectés au réseau (GCPV). Pour cela, deux BPNN sont nécessaires : le premier ANN est nécessaire pour la classification du courant à partir de deux données d'entrée (niveau d'irradiation solaire et courant au point de puissance maximal), tandis que le deuxième ANN est appelé pour la classification de tension à partir des deux données d'entrée (température de la cellule et tension au point de puissance maximal). La sortie des deux ANNs représente l'entrée de l'algorithme combinatoire afin d'obtenir le diagnostic du générateur PV caractérisant l'étape finale de cette approche. Cet algorithme diagnostique les défauts les plus fréquents rencontrés dans les installations PV qui sont : un module en court-circuit dans le générateur PV, deux modules en court-circuit dans le générateur PV, quatre modules en court-circuit dans les générateurs PV et une chaîne déconnectée dans le générateur PV. Les résultats obtenus de la méthode citée sont excellents avec une précision globale moyenne de 98,6 %.

Afin de trouver le meilleur choix d'ANNs, cinq types d'algorithmes ont été testés, à savoir : le réseau neuronal à propagation arrière (BPNN), le réseau neuronal probabiliste (PNN), le réseau neuronal à régression généralisée (GRNN) et deux réseaux neuronaux à fonction de base radiale. (RBF). Ces ANN ont été testés et comparés en utilisant les mêmes défauts et les mêmes conditions de travail. Les résultats obtenus ont apporté une bonne clarification et démontrent que l'algorithme PNN prend la tête de liste comme le meilleur ANN du point de vue de son temps de réponse ainsi que de son affichage de 100% dans tous les concepts statistiques par rapport à d'autres algorithmes.

L'efficacité de la méthode développée est évaluée expérimentalement en utilisant des données réelles mesurées, collectées à partir d'un système PV connecté au réseau GCPV réel situé au Centre Des Energies Renouvelables à Alger (Algérie).

Mots clés: installation photovoltaïque connectée au réseau, détection des défauts, diagnostic, Réseau de neurones artificiels (ANN), conditions de travail, point de puissance maximale (MPPT).

الملخص

في هذه الأطروحة، تم تطوير طريقة جديدة لتشخيص الأنظمة الكهروضوئية باستخدام الذكاء الاصطناعي. تستخدم طريقة التشخيص هذه مصنفات الشبكات العصبية الاصطناعية ذات الانتشار الخلفي (BPNN) لإكتشاف وتشخيص الأعطال في جانب التيار المباشر (DC) لأنظمة الخلايا الكهروضوئية المتصلة بالشبكة. (GCPV) لهذا ، يلزم وجود شبكتين: BPNN أولANN ضروري لتصنيف التيار انطلاقا من معطيات إدخال إثنين (مستوى التشعيع الشمسي والتيار عند نقطة الطاقة القصوى) ، بينما ثاني ANN ضروري لتصنيف الجهد انطلاقا من بيانات إدخال إثنين (الخلية درجة الحرارة والجهد عند نقطة الطاقة القصوى). يمثل ناتج كل من ANNs مدخلات الخوارزمية التوافقية من أجل الحصول على تشخيص للمولد الكهروضوئي الذي يميز الخطوة الأخيرة من هذا النهج. تقوم هذه الخوارزمية بتشخيص الأخطاء الأكثر شيوعًا في التركيبات الكهروضوئية وهي: وحدة واحدة قصيرة الدائرة في المولد الكهروضوئي ، ووحدتان قصيرتان في المولد الكهروضوئي ، وأربع وحدات قصيرة الدائرة في المولدات الكهروضوئية وسلسلة واحدة غير متصلة في المولد الكهروضوئي. النتائج التي تم الحصول عليها من الطريقة المذكورة ممتازة بمتوسط دقة إجمالية 98.6٪.

من أجل العثور على أفضل خيار لشبكاتANN ، تم إختبار خمسة أنواع من الخوارزميات وهي: الشبكة العصبية للانتشار الخلفي (BPNN)، والشبكة العصبية الاحتمالية (PNN) ، والشبكة العصبية المعممة للانحدار (GRNN) وإثنين من الشبكات العصبية للوظيفة الشعاعية . (RBF) لقد استخدمت هذه الشبكات العصبية الاصطناعية نفس الأخطاء ونفس ظروف العمل. قدمت النتائج التي تم الحصول عليها توضيحًا جيدًا وأظهرت أن خوارزمية PNN تتصدر القائمة كأفضل ANNمن وجهة نظر وقت استجابتها وكذلك عرضها 100٪ في جميع المفاهيم الإحصائية مقارنة بالخوارزميات الأخرى.

يتم تقييم كفاءة الطريقة المطورة بشكل تجريبي باستخدام بيانات مقاسة حقيقية، تم جمعها من نظامGCPV الحقيقي الموجود في مركز الطاقات المتجددة في الجزائر العاصمة (الجزائر)

الكلمات الرئيسية: النظام الكهروضوئي المتصل بالشبكة ، وإكتشاف الأعطال، والتشخيص ، الشبكة العصبية الاصطناعية (ANN) ، ظروف العمل ، نقطة الطاقة القصوى .(MPPT)

LIST OF ABBREVIATIONS

AC	Alternative Current.
AI	Artificial Intelligence.
ANNs	Artificial Neural Networks.
bi	Bias of Back Propagation Neural Network.
BPNN	Back Propagation Neural Network.
DC	Direct Current.
ECM	Earth Capacitance Measurement.
EWMA	Exponentially Weighted Moving Average Estimation.
FF	Fill Factor.
FL	Fuzzy Logic.
FN	False negative, signifies that the samples contain characteristics of a specific class and indeed they are not classified in this class
FP	False positive, signifies that the samples do not contain characteristics of a specific class and they are classified in this class.
G	Irradiance level.
GA	Genetic Algorithms.
GCPV	Grid Connected Photovoltaic Systems.
GRNN	Generalized Regression Neural Network.
IFD	Intelligent Fault Diagnosis.
IV_f	I-V Curve Inflexion Factor.
k	Boltzman Constant (1.3806503 x 10-23 J/°K).
kNN	k-Nearest Neighbors.
MAPE	Mean Absolute Percentage Error.
MLP	Multi-Layer Perceptron.
MPP_f	Maximum Power Point Factor.
MPPT	Maximal Power Point Tracking.
n	Diode Ideality Factor.
N	Size of the Database.
PDF	Probability Density Function in Probabilistic Neural Network.
P_{mpp}	Power at Maximal Power Point.
PN	Positive, Negative junction.

PNN	Probabilistic Neural Network.
PS	Partial Shading.
PSO	Particle Swarm Optimization.
PV	Photovoltaic Systems.
q	Electron Charge (1.6 0217646 x 10-19 C).
R^2	Coefficient of Determination.
RBF	Radial Basis Function Neural Network.
RC	Ratio Current.
RMSE	Root Mean Square Error.
TDR	Time Domain Reflectometry.
TN	True negative, signifies that the samples do not contain characteristics of a specific class and indeed they are not classified in this class.
TP	True positive, signifies that the samples contain characteristics of a specific class and indeed they are classified in this class.

LIST OF SYMBOLS

ωi	Weight Vector of RBF.
σ	Standard Deviation.
$\hat{C}(x)$	Final electrical parameter classification.
D	Denominator in Generalized Regression Neural Network.
d_{max}	Maximum distance between the cluster centers selected in Radial Basis Function.
E_i	Current Error.
EL_{c_ref}	Error Between Measured And Simulated Capture Losses Of The Healthy System.
E_V	Voltage Error.
I_0	Diode Saturation Current.
I_{ac}	Alternative Current.
I_{cc}	Short Circuit Current.
I_{mpp}	Current at Maximal Power Point.
I_{ph}	Photo Generated Current.
L_{cm}	Miscellaneous Capture Losses.
L_{ct}	Thermal Capture Losses.
N	Numerator in GRNN.
R_s	Serie Resistance.
R_{se}	Equivalent Series Resistance.
R_{sh}	Shunt Resistance (parallèle).
RV	Ratio Voltage.
T_c	PV Cell or Module Temperature.
Th	Threshold between measured and simulated data.
V_{ac}	Alternative Voltage.
V_{mpp}	Voltage at Maximal Power Point.
V_{oc}	Open Circuit Voltage.
W_i	Weight of Back Propagation Neural Network.
y_n	n^{th} Measured Data.
$\hat{y}_n$	n^{th} Simulated Data.
σ_1	Spread of Radial Basis Function.
φ	Activation function of Radial Basis Function.
φi	Sigmoid Function of Back Propagation Neural Network.

TABLE OF CONTENTS

DEDICATION I

ACKNOWLEDGEMENTS II

ABSTRACT III

RESUME IV

الملخص V

LIST OF ABBREVIATIONS VI

LIST OF SYMBOLS VII

TABLE OF CONTENTS VIII

LISTE OF ILLUSTRATIONS AND GRAPHICS IX

LISTE OF TABLES X

INTRODUCTION 1

CHAPTER 1 STAT OF ART OF FAULT DETECTION AND DIAGNOSIS OF GRID CONNECTED PHOTOVOLTAIC SYSTEMS 4

1.1. Introduction 4
1.2. The various faults encountered in PV systems 5
1.2.1. External problems 5
1.2.2. Internal problem 9
1.3. The different categories of fault detection and diagnosis in PV systems 11
1.3.1. Model based diagnosis technique 11
1.3.2. Signal processing based fault diagnosis methods 13
1.3.3. Artificial intelligence in fault diagnosis techniques 15
1.4. Conclusion 20

CHAPTER 2 MODELLING AND SIMULATION OF PHOTOVOLTAIC GENERATOR 21

2.1. Introduction 21
2.2. Methodology 21
2.3. Presentation of experimental PV plant 22
2.4. PV array modelling: 26
2.4.1. Modeling of PV cell 26
2.4.2. PV module / array modelling and I-V characteristics: 27
2.5. PV array model simulation: 30
2.5.1. PV array data validation: 32
2.5.2. Healthy real sub array versus healthy simulated sub array 34
2.5.3. Current validation of maximum power point. 36
2.5.4. Voltage validation of maximum power point. 38
2.5.5. Faulty sub array versus healthy sub array 40

2.5.6. Mathematical analysis of faulty electrical parameters 44
2.6. Conclusion 49

CHAPTER 3 FAULT DIAGNOSIS OF PHOTOVOLTAIC GENERATOR USING BACK PROPAGATION NEURAL NETWORK. 50

3.1. Introduction 50
3.2. Methodology 50
3.3. IFD algorithm's description 51
3.3.1. The various faults treated in this study 51
3.3.2. The ANNs architecture 54
3.4. Diagnosis procedure in PV generator 55
3.5. Simulation Results and interpretations 60
3.5.1. Behavior of ANNs in PV generator fault diagnosis 60
3.5.2. Global test Classification 68
3.5.3. Performance of the system 70
3.6. Conclusion 71

CHAPTER 4 COMPARATIVE STUDY OF PV GENERATOR DIAGNOSIS WITH FOUR DIFFERENT NEURAL NETWORKS. 72

4.1. Introduction 72
4.2. Methodology 72
4.3. Comparative study 73
4.3.1. Phase 1: ANNs choice 73
4.3.2. Phase 2: Effectiveness evaluation 81
4.4. Results presentation and discussions 83
4.4.1. BPNN classification 87
4.4.2. RBF classification 89
4.4.3. PNN classification 93
4.4.4. GRNN classification 94
4.4.5. Training phase and respond time 96
4.4.6. Analysis the efficiency of the five different algorithms 99
4.5. Conclusion 107

General Conclusion 108

REFERENCES 111

APPENDIX 'A' 118

APPENDIX 'B' 138

LISTE OF ILLUSTRATIONS AND GRAPHICS

CHAPTER 1

Figure 1.1. Faults encountered in PV systems……. 5
Figure 1.2. Soiling on a PV Generator……... 6
Figure 1.3. Operation of PV array (a) under uniform insolation (b) under partial shading (c) the resulting *I–V* and *P–V* curve for (a) and (b)……. 7
Figure 1.4. Snow on a PV Generator……. 7
Figure 1.5. Tree leaves and bird droppings on a PV Generator……. 8
Figure 1.6. Power outage from the grid……. 8
Figure 1.7. Branch disconnected in PV Generator…….. 9
Figure 1.8. Degradation of PV Cells…….. 10

CHAPTER 2

Figure 2.1. Scheme of PV model parameters identification…….. 22
Figure 2.2. Roof grid connected PV plant in Algiers, Algeria……. 23
Figure 2.3. Experimental setup……. 24
Figure 2.4. Daily profile of onsite measured (a) Irradiance, (b) Temperature, (c) Output power……. 25
Figure 2.5. Equivalent circuit of solar cell…….. 26
Figure 2.6. I-V and P-V curves…… 28
Figure 2.7. Effect of Solar Irradiation (a) I-V curves ……. 28
Figure 2.7. Effect of Solar Irradiation (b) P-V curves (the following)……. 29
Figure 2.8. Effect of Cell Temperature (a) I-V curve……. 29
Figure 2.8. Effect of Cell Temperature (b) P-V curves……. 30
Figure 2.9. Block diagram of the studied PV sub-array in Simscape /MATLAB……. 31
Figure 2.10. Experimental validation data (a) (b)……. 32
Figure 2.10. Experimental validation data (c) (d)……. 33
Figure 2.11 Measured and simulated PV module output data……. 35
Figure 2.12. Current of real measured data against simulated data of PV generator……. 36

Figure 2.13. Part of figures 2.12 zoomed……. 37
Figure 2.14. Absolute error between measured and simulated current at MPP……. 37
Figure 2.15. Identification of the current of maximum power point in normal operation conditions……. 38
Figure 2.16. Voltage Measured real data against simulated data of PV generator…….. 38
Figure 2.17. Part of figure 2.21 zoomed……. 39
Figure 2.18. Absolute error between measured and simulated voltage at MPP…….. 39
Figure 2.19. Identification of the maximum power point voltage in normal operation conditions……. 40
Figure 2.20. Example of temporary PV faults……. 41
Figure 2.21. Various examples of shading PV faults (the following) (a)…… 41
Figure 2.21. Various examples of shading PV faults (the following) (b)…… 42
Figure 2.22. Examples of permanents PV faults……. 43
Figure 2.23. The impact of the string fault on the current……. 44
Figure 2.24. Faulty simulated current via healthy measured current of the system……. 45
Figure 2.25. The voltage of a PVG with one short-circuited module……. 46
Figure 2.26. The voltage of a PVG with two short-circuited modules……. 46
Figure 2.27. The voltage of a PVG with four short-circuited modules……. 47
Figure 2.28. The voltage of global different short-circuited modules in PVG 47
Figure 2.29. Faulty simulated voltage via healthy measured voltage of the system……. 48

CHAPTER 3

Figure 3.1. Schematic description of the IFD methodology…….. 51
Figure 3.2. Overall scheme of the PV system (a) normal operation conditions (b) faulty string and short circuits faults (1)……. 52
Figure 3.2. Overall scheme of the PV system (a) normal operation conditions (b) faulty string and short circuits faults (2)……. 53
Figure 3.3. Architecture of ANN current…….. 54

Figure 3.4. Architecture of ANN current…….. 55
Figure 3.5. Flowchart of the IFD algorithm…… 57
Figure 3.6. ANN's training phase of Current…….. 60
Figure 3.7. ANN's training Error of Current…… 61
Figure 3.8. Classification accuracy of Voltage…….. 61
Figure 3.9. Impp classification using ANN…… 62
Figure 3.10. ANN's training phase of Voltage…… 63
Figure 3.11. ANN's training Error of Voltage…… 64
Figure 3.12. Classification accuracy of Voltage…… 64
Figure 3.13. Vmpp classification using ANN…….. 65
Figure 3.14. Global diagnosis of the system…….. 66
Figure 3.15. IFD approach test results…… 68

CHAPTER 4

Figure 4.1. Architecture of BPNN model of the present work…….. 72
Figure 4.2. Architecture of RBF model of the present work…… 74
Figure 4.3. Architecture of PNN model of the present work…… 77
Figure 4.4. Architecture of GRNN model of the present work…… 79
Figure 4.5. Impp classification using BPNN…… 85
Figure 4.6. Vmpp classification using BPNN…….. 86
Figure 4.7. Global diagnosis of the system using BPNN…… 86
Figure 4.8. Impp classification using RBF1…… 87
Figure 4.9. Vmpp classification using RBF1…….. 88
Figure 4.10. Global diagnosis of the system using RBF1…… 88
Figure 4.11. Impp classification using RBF2…… 89
Figure 4.12. Vmpp classification using RBF2…… 90
Figure 4.13. Global diagnosis of the system using RBF2…… 90
Figure 4.14. Impp classification using PNN…… 91
Figure 4.15. Vmpp classification using PNN…….. 91
Figure 4.16. Global diagnosis of the system using PNN…… 92
Figure 4.17. Impp classification using GRNN…… 93
Figure 4.18. Vmpp classification using GRNN…….. 93
Figure 4.19. Global diagnosis of the system using GRNN…….. 94

Figure 4.20. Confusion of the global accuracy in the five different algorithms……. 99
Figure 4.21. Confusion of the global sensivity in the five different algorithms…… 100
Figure 4.22. Confusion of the global precision in the five different algorithms…….. 102
Figure 4.23. Confusion of the global specificity in the five different algorithms…….. 103
Figure 4.24. Confusion of the global criterions in the five different algorithms…….. 104

LISTE OF TABLES

CHAPTER 1

Table 1.1. Various type of fault…… 16
Table 1.2. Different type of faults occurring in the examined PV plant…… 18

CHAPTER 2

Table 2.1. Electrical properties of the Isofoton 106-12 PV module (STD: solar irradiation=1000W/m2, cell temperature= 25°C)…….. 23
Table 2.2. Model performance metrics values…… 36
Table 2.3. Faults correspondence table…… 48

CHAPTER 3

Table 3.1. The different state of the system with faults and their symbols…….. 54
Table 3.2. Classifications stage, classes and symbols……... 56
Table 3.3. Different combination of classes obtained…… 56
Table 3.4. Faults correspondence table…….. 58
Table 3.5. Faults correspondence table…… 59
Table 3.6. Faults correspondence table for ANN test phase…… 67
Table 3.7. Faults correspondence table for ANN test phase…… 68
Table 3.8. Performance of classification with ANN (%)…….. 69

CHAPTER 4

Table 4.1. Confusion matrix under intermittent classification troubles……. 81
Table 4.2. Faults correspondence table of current…… 83
Table 4.3. Faults correspondence table of voltage…… 84
Table 4.4. Parameters adjustable of each neural network…… 95
Table 4.4. Parameters adjustable of each neural network (the following)…… 96
Table 4.5. The variation of running time for the five algorithms…….. 97
Table 4.6. The nomination of the five algorithms using IFD…… 97

Table 4.7. Current Accuracy (%)…….. 97
Table 4.8. Voltage Accuracy (%)…….. 98
Table 4.9. Global Accuracy (%)…….. 98
Table 4.10. Current sensivity (%)…….. 99
Table 4.11. Voltage Sensivity (%)…….. 100
Table 4.12. Global Sensivity (%)…….. 100
Table 4.13. Current Precision (%)…….. 101
Table 4.14. Voltage Precision (%)…….. 101
Table 4.15. Global Precision (%)…….. 101
Table 4.16. Current Specificity (%)……... 102
Table 4.17. Voltage Specificity (%)…….. 103
Table 4.18. Global Specificiy (%)…….. 103
Table 4.19. Localization of confusions in all algorithms…….. 105

INTRODUCTION

Nowadays, modern civilization is looking for a profound and global energy change throughout the world, from fossil fuels such as natural gas, oil, lignite and coal resources towards renewable energies. The goal of this change is to avoid catastrophic climate change that affects the health's deterioration of the current and future human generations as well as for countless other species. However, reliance on renewable energies has become an inescapable trend as it plays a key role in decarbonizing the global energy system for decades to come. In addition, the production of renewable energy exceptionally photovoltaic energy growing up in staggering rate for its cleanliness, safety, quiet, reduction in energy bills and low maintenance requirement, it will be the principal source of energy in the world in the coming years increasing from 330 TWh in 2019 to almost 3300 TWh in 2030, according to reference [1].

keen to initiate and succeed its energy transition, Algeria has adopted an ambisous plan to develop and promote renewable energies As reported by Shariket Kahraba wa Taket Moutadjadida (SKTM), which is a subsidiary of the Algerian company of electricity and gas (SONELGAZ), Algeria is massively engaged in renewable energy especially Grid Connected PV systems with 23 grid connected PV plants using poly crystalline PV modules fixed in: highlands (east: Batna, Souk Ahras, Setif, Bourdj Bouararidj and Mila; west: Sidi Belabes, Saida, Nâama et El Bayodh), center of the country (M'sila, Djalfa, Laghouat and Ouargla), the south of the country (In Salah, Adrar and Timimoun) for a total power of approximately 330 MWc [2].

To achieve the best power-generation efficiency, PV systems must work under particular conditions, unshaded area, high irradiation level, low temperature and PV panel optimal orientation. In addition, PV panels must be clean enough because the accumulation of dirt, dust, sand as well as snow mask the solar irradiation and reduce the efficiency of the global system. For this purpose, the techniques of regular maintenance and detection become more and more crucial. In order to assure a good power production, reliability, efficiency, safety and quality in global PV systems, fault detection and diagnosis become necessary and have shown exceptional interest last few times. Nowadays, real-time fault detection and diagnosis in PV systems is catching the eye

of a large number of researchers in this field, in which various research studies and investigations have been discussed [3].

The analysis of the developed techniques has made it possible to identify three different categories. The first category contains diagnostic techniques, which are based on the system's model in order to identify the appearance of faults [4]. The simulation data are compared to the system outputs, where the inputs of the model are particularly the weather parameters and electrical parameters of the installed PV array [5-12] and the output of the model consists of the maximum power point of the generator. Supplementary parameters such as AC input/output ratio, DC input/output ratio and reference yield have been used [13-19]. The second category includes mathematical or statistical analysis-based methods such as time domain reflectometry and Fourier analysis [20,21]. Earth capacitance measurement diagnostic method [22, 23] as well as the stacked auto-encoder and clustering [24], non-parametric kernel density estimation method [25] and learning algorithm [26]. These techniques are based on several information which stems generally from (I-V) such as curve open circuit voltage, short circuit current, ideal factor, fill factor indicator, values of shunt and series resistors in a PV module, a PV string or branch and a PV Generator [27-31]. Another method for these category is based on the analysis of three predefined indicators [32]. The third category is focused on the computational intelligence-based techniques classifier, including, Artificial Neural Networks (ANNs), Fuzzy Logic (FL), Genetic Algorithms (GA) and k-Nearest Neighbors (kNN) [33-56].

In accordance with the above presented arguments belonging to different categories, an original intelligent fault detection and diagnosis method for grid connect PV systems especially in PV generator is proposed in the present thesis. The main idea consists to use the Back Propagation Neural Networks, which are identified by their efficient learning capacities, generalization and good classification applying with excellent results [48] to detect and then identify faults for grid connected PV plant in DC part. After this, a comparison study is performed to explore the influence of the ANNs choice on the Diagnosis' performance and efficiency. For this, five different ANNs have been

considered: Back Propagation Neural Networks, Probabilistic Neural Networks, Generalized Regression Neural Network and two algorithms of Radial Basis Function Neural Networks [57].

The thesis is organized into four chapters, where the last three chapters present the exclusive works, which have been published.

The first chapter presents the problematic related to faults encountered in PV installations. Then, a literature review on the already proposed methods of faults detection and diagnosis is conducted.

In the second chapter, are presented the experimental setup as well as the developed model and the results of its simulation and experimental validation. This model has been developed using Simscape MATLAB/ simulink , followed by the elaboration of pertinent database including healthy and faulty operations and finally an analytical study presents the faulty behavior based on the deviation between the desired output and the experimental real output.

The third chapter is devoted in detail to the application of BPNN classifiers for faults detection and diagnosis. The architecture of both ANN classifying the current and the voltage at the maximum power point, Impp and Vmpp, are present, in addition to diagnosis procedure in PV generator and finally, the training, validation phase of both ANNs.

The last chapter provides a comparative study and analyses the impact of the Artificial Neural Network choice on the diagnosis performances using four major criterions: accuracy, sensivity, specificity and precision. To achieve this goal, five ANNs are studied: Back-Propagation Neural Network (BPNN), Probabilistic Neural Network (PNN), Generalized Regression Neural Network (GRNN) and two Radial Basis Function Neural Network (RBF1, RBF2).

And finally, a general conclusion summarizes the research results obtained within the framework of this thesis and the proposed perspectives.

CHAPTER 1 STAT OF ART OF FAULT DETECTION AND DIAGNOSIS OF GRID CONNECTED PHOTOVOLTAIC SYSTEMS

1.1. Introduction.

Up until now, Renewable energy is one of the biggest and most significant innovations of the century. The reason is that most of renewable energies have specific benefits promoting protection of the biodiversity. In fact, these types of energies are considered as clean energies by their planet's protection from possible climatic upheaval. Additionally, renewable energies are a solution to affordable, clean, healthy, and eco-friendly power contrary to classical energies coming from fossil fuels such as: petroleum, natural gas, coal or materials like uranium which are limited on the planet and on the other side pose a huge problem at ecosystems by emitting a lot of greenhouse effect which contribute to global warming as well as carbon di-oxide emission and fine particle pollution [58]. Moreover, the renewable energy's production growths exponentially especially photovoltaic energy that plays a key role in the electricity production in the future [4].

To achieve the best power-generation efficiency, PV systems must work under particular conditions; unshaded area, high irradiation level, low temperature and PV panel optimal orientation. In addition, PV panels must be clean as the accumulation of dirt masks the solar irradiation and reduces the efficiency of the global system. This necessitates regular maintenance in addition to the monitoring. In fact, during their long lifetime, of about 25 years, PV systems can be the subject of numerous faults. That explains the importance of the diagnosis and fault detection for PV systems which is necessary not only to increase system power generation reliability but also for operating costs reduction. Real-time diagnosis of PV system has drawn many researchers attention nowadays. As a result, a great deal of recent research on PV systems have been focusing on that area to help to make possible the fault detection and isolation especially for the PV generator.

1.2. The various faults encountered in PV systems.

One of the main research objectives on grid connected PV systems is to improve the efficiency, availability and reliability of the systems. Two principal problems exist in grid connected PV systems can influence on its performance: external problems that are not consider in this research and internal problem that is addressed in our study. However, if these problems are not detected and localized can not only reduce the production of electricity and improve the hot spots but also threaten the availability, reliability and effectively the security of the global system.

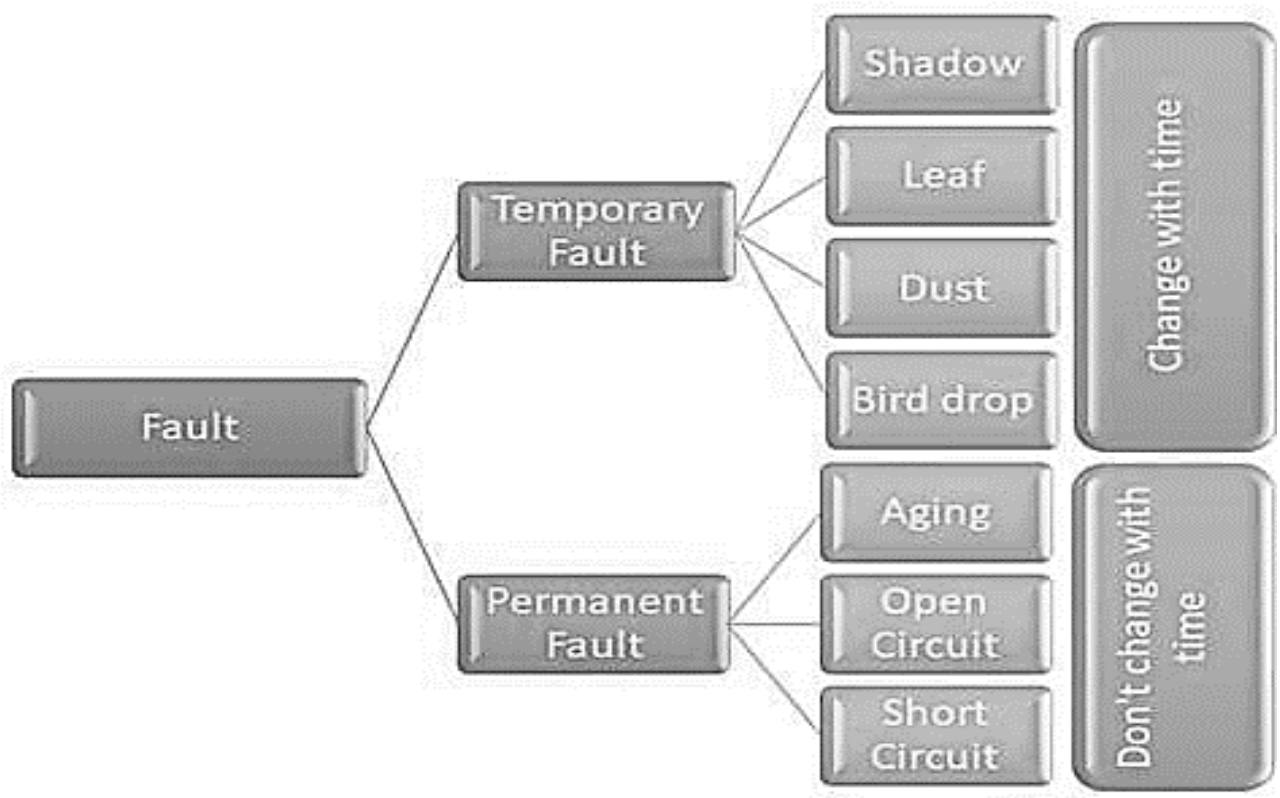


Figure 1.1. Faults encountered in PV systems.

1.2.1. External problems

1.2.1.1. The soiling

The dirt has an important influence on the performances of PV systems, particularly in arid regions. The accumulation of dirt on PV array contribute to reduce the normal current generation capacity of a PV module, a PV branch or PV generator.

Consequently, the maintaining a clean panel surface is important in order to avoid energy loss.

Figure 1.2. Soiling on a PV Generator.

1.2.1.2. Shading

In general, the shading occurs when the buildings, the objects and other things such as trees is located close the sun and PV generator. This represents one of the biggest sources of energy loss in the operation of the PV system. The shading can: contribute to reduce the output power, cause the thermal constraint on the PV modules due to high cell temperature and even reverse voltage of shaded cells due to overheating. In order to overcome the influences of shading and minimize energy loss, the bypass diodes are wired in parallel to a number of solar cells to continue to conduct current around the shaded cells or modules, while they are blocked under normal operating conditions.

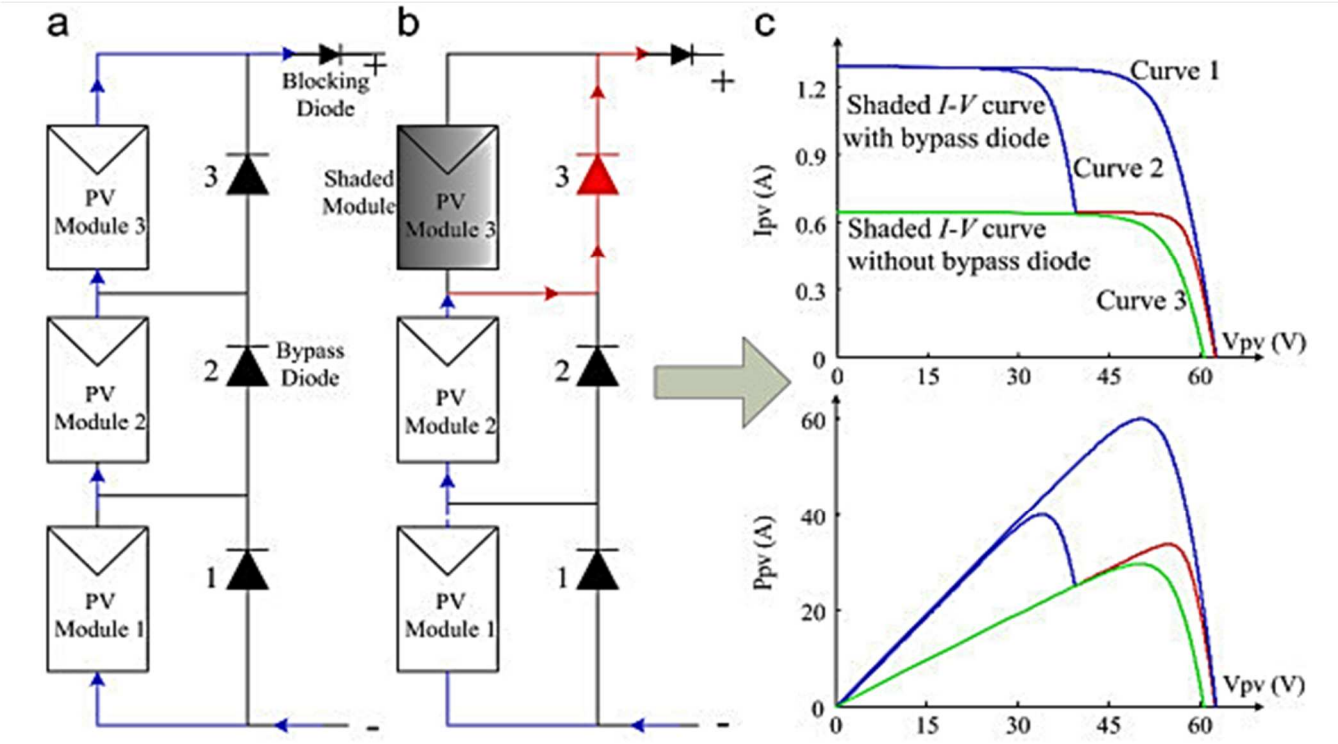


Figure 1.3. Operation of PV array (a) under uniform insolation (b) under partial shading (c) the resulting *I*–*V* and *P*–*V* curve for (a) and (b) [59].

1.2.1.3. Snow cover

The snow cover only occurs on cold winter days. Snowfall has a negative impact on performance of the PV system such as the reduction of output power. In addition, if one PV module is blocked by snow, the whole branch will perform poorly, even if there is a high global radiation available.

Figure 1.4. Snow on a PV Generator.

1.2.1.4. Tree leaves and bird droppings

Due to its exposed surface, tree leaves and bird droppings can fall on PV panels, which can contribute to reduce the energy production of some cells. Bird droppings and tree leaves will cause a much crucial loss of power than dirt and dust. Therefore, to avoid this problem, the cleaning and prompt removal of these waste are indispensable for the PV panels. In these cases, the by-pass diodes are necessary to reduce the important effect on PV systems cited in paragraph above.

Figure 1.5. Tree leaves and bird droppings on a PV Generator.

1.2.1.5. Power cut from electrical network

The production of electricity is highly dependent on weather conditions and time of day. If an accident or natural disaster occurs and the grid power is lost, the power supply to the PV generator will also be cut off. A power cut from the grid does not mean any energy generation for the PV system during this period.

Figure 1.6. Power outage from the grid.

1.2.1.6. Faults with total failure in PV systems

The total failure is defined as a permanent loss of power during pleasant weather days. If there is no data sent from the PV system, failures due to the total blackout can be assumed. Faults in PV systems not only decrease the net performance of these systems, but can also threaten the safety and reliability of the entire system. Generally, faults due to total failure are easily diagnosed. However, other faults of PV system are not easily detectable due to uncertainties on the overall efficiency values and the absence of reference values.

1.2.2. Internal problem

Usually, the PV generator owns many branches in parallel and the output of each branch can be connected to junction box. Knowing that the branch can meet a fault such as the disconnection or the degradation. The principal difference between both faults is that the disconnection can be defined as a constant loss of energy contrary to the degradation that is considered as a changing loss of energy.

1.2.2.1. Disconnection

If a branch has a disconnect problem, the output of this string is zero and the energy loss is constant. Under these circumstances based on the percentage of energy loss, the number of disconnected branches can be detected at the location of the inverter.

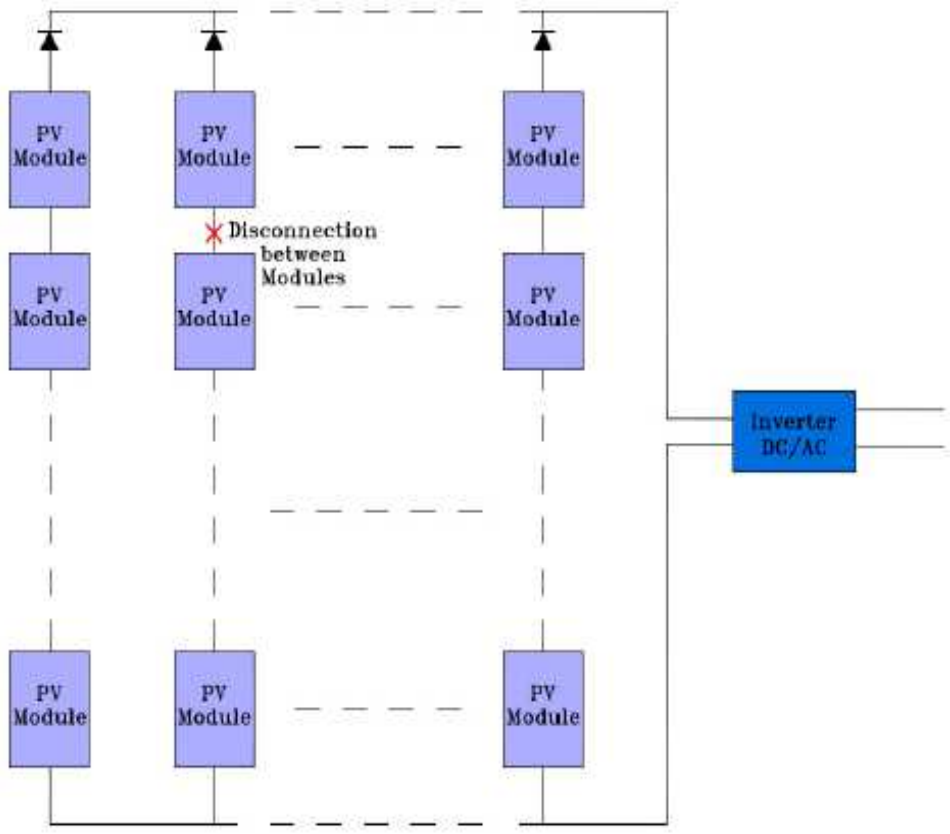


Figure 1.7. Branch disconnected in PV Generator.

1.2.2.2. Degradation

Power degradation resulting from aging solar cells. It plays a key role in decreasing of the output power during its lifetime and it differs from technology to another. The reason could be the increase in series resistance between modules due to decrease in contact adhesion or corrosion caused by water vapor or reduction in shunt resistance connected parallel to junction PN. Additionally, the loss of the anti-reflective coating could also lead to degradation of the power. Degradation not only contributes to declining PV systems performance, such as reducing the output power of a PV module, but can also lead to cell damage and early system failure due to corrosion [60]. If the degradation occurs in a chain, the I-V curve might be distorted, but the branch would continue to produce power.

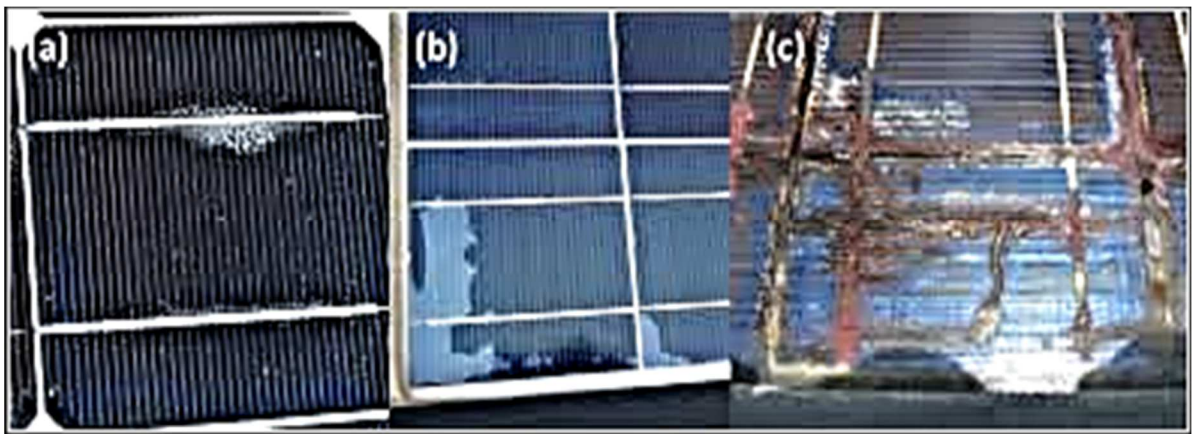


Figure 1.8. Degradation of PV Cells.

Therefore, developing a method for detecting faults in a PV system is crucial to minimize energy loss and improve the performance of the PV system. Several techniques have been proposed for the detection and diagnosis of faults in solar photovoltaic systems representing in the next section.

1.3. The different categories of fault detection and diagnosis in PV systems.

On the basis of a literature overview, it was concluded that PV diagnosis techniques can be classified in three number of categories: Model based diagnosis technique, Signal processing based fault diagnosis methods and Artificial intelligence in fault diagnosis techniques. These methods are based on many crucial between weather and electrical parameters such as: cell temperature , solar Irradiation , current and voltage at Maximal Power Point (I_{mpp}, V_{mpp}), short circuit current I_{sc}, open circuit voltage V_{oc}, the output of inverter parameter such as alternative current and voltage, I_{ac} and V_{ac} respectively, as well as the current-voltage (I-V) curve characteristics.

1.3.1. Model based diagnosis technique

This category includes model based diagnosis techniques [5] in which the model of the system is used to decide about the occurrence of faults. For this, the simulation values are compared to the system outputs. The inputs of the model are mainly the

meteorological working conditions and the electrical parameters of the installed PV modules [7].

An efficacious method of fault detection and diagnosis in GCPV systems has been proposed by Mahmoud Dhimish et al in [6] based on voltage and power ratios. This approach started by the simulation of the theoretical performance of the PV system in goal to compute the theoretical outputs voltage and power. After that, the ratios between the theoretical and measured voltage and power are calculated and analyzed, in order to detect the fault occurrence and diagnose its type. This method is entirely dependent on the power and voltage ratios, although she is recognized by its high capacity for fault detection and diagnosis. Consequently, it depends on the accuracy of both the simulation model and the measuring devices.

In reference [7], Chouder et al have developed an automatic supervision and fault detection procedure for PV systems based on the power losses analysis. The principal objective of this approach consists to include parameter extraction techniques to calculate main PV system parameters from monitoring data in real conditions of work, taking into account the environmental irradiance and module temperature evolution, allowing simulation of the PV system behavior in real time. The system is considered under faulty operation if the measured power losses are beyond these boundaries. Finally, in order to diagnose the fault type, current and voltage ratios are evaluated and monitored. Two novel indicators of power losses have been defined in this work: Thermal capture losses (Lct) and Miscellaneous capture losses (Lcm). In addition, two another indicators of the deviation of the DC variables respect to the simulated ones have been also defined. These indicators are the current and voltage ratios: *RC* and *RV.* The global indicators are used to diagnose the fault types. This method has approved an efficient result tested for three operating cases: healthy system operation, faulty string operation and operation in the presence of partial shadowing, aging and MPPT error.

On the other part, a procedure for automatic supervision, fault detection and diagnosis for different faults of Grid connected PV systems in real time has been proposed in [8]. A comparative study between measured and simulated yield has been

achieved to detect faults in the PV plan, while the fault diagnosis is carried out by analyzing and comparing DC current and DC voltage with a set of healthy system thresholds. This approach needs the measured meteorological conditions and the electrical parameters to develop a LabVIEW-based software of: firstly capture losses computations, secondly PV system simulation, thirdly online parameters supervision, and finally fault detection and diagnosis [10]. EL_{c_ref} is the error between measured and simulated capture losses of the healthy system σ is standard deviation of this error. The regular check of this error is required to detect any fault occurrence; i.e. if its value does not exceed predefined thresholds, the system is considered working under healthy operation. When the presence of faults is detected, two novels indicators, called current error E_i and voltage error E_V are computed. In order to process to diagnose the detected fault, a set of predefined thresholds for E_i and E_v should be computed. This method has been experimentally tested under an actual GCPV system to detect and diagnose the occurrence of: partial shading, ground fault, short circuit fault of PV modules, short circuit of bypass diode and string disconnection in the array.

Additional parameters have been used while dealing with grid connected PV power plants such as DC input/output ratio, AC input/output ratio and reference yield measurements [14-15]. Model-based fault detection methods use residuals which indicate changes between the PV generator and its model. This difference is used to estimate the increase of the series resistance and the decrease of the parallel resistance using fill factor method [16] and can be used to focus on intra-string line-line fault [17] or to detect the shading using exponentially weighted moving average estimation (EWMA) [18]. Another method to detect and diagnose faults such: short circuit, open circuit, shading and degradation in PV arrays by residuals using current-voltage c6urves and ambient conditions based on deep neural network has been also developed [20].

1.3.2. Signal processing based fault diagnosis methods

This type of category is classified the signal processing based fault diagnosis methods. This includes methods based on mathematical or statistical analysis such as time domain reflectometry [21]. In this approach, the time domain reflectometry (TDR) applies the voltage signal into the string and observes the signal response waveforms, was applied to the PV strings containing faults to detect the fault and locate its position.

The results showed the disconnections and the degradations with series resistance increase between modules were detected with the response signal voltage rise and the signal rise timing shift.

Moreover, these techniques should be combined to other diagnostic methods as earth capacitance measurement to expand their faults diagnosis capability [22, 23], The signal shift from the inputted signal to the reflected signal is translated into the fault position in the line, and the waveform change is translated into the mismatch type (open circuit, short circuit, resistance increase, etc.) and the amount of the fault (impedance change). Compared to ECM, which detects only the disconnection in the string, TDR is more promising because it can detect not only the disconnection but also the impedance changes with degradation. Moreover, TDR has been used to detect and localize degradation fault.

Furthermore, Measurements of the current-voltage (I-V) curve are one method of PV fault detection. This curve can provide crucial information such as short-circuit current, open circuit voltage, fill factor indicator, ideal factor, series and shunt resistors values for a PV module, a PVG, a string or branch or PV field [27, 28]. These methods are known to be very efficient to detect and isolate the presence of shading and soiling of the PV module, of the degradation of the solar cells and balance of system components. Thus, this method consists mainly on three stages.

- The first one relies on measuring the (I-V) curve and the in-plane irradiance (G) of the tested PV generator (module, string or array). The in-plane irradiance level (G) could be obtained by using irradiance sensors or mathematical estimations.

- In the second stage, the in-plane irradiance level (G) is analyzed, such that if its value is lower than 500 W/m², the diagnosis procedure will be stopped.
  The main reason of this limitation is that the power losses, due to the shading and the increase in series resistance, cannot be visible, using I-V curves under low irradiance level [28]. After that, if the in-plane irradiance is higher than 500 W/m², the measured I-V characteristic will be filtered from any noise and used to compute a

set of diagnostic indicators. The indicators are then normalized to maintain the independency with the system configuration.

- Finally, the normalized indicators of diagnosis are analyzed automatically, based on three fuzzy logic classifiers, to detect and diagnose the faults mentioned above [28]. The indicators that have been used in this method are: the I-V curve inflexion factor IV_f, the maximum power point factor MPP_f, the equivalent series resistance R_{se} and the fill factor FF.

This method can only be applied for the string inverter topology, which is not the most frequently used topology nowadays, and it gives efficient results only under high irradiance level ($G>500W/m^2$).

1.3.3. Artificial intelligence in fault diagnosis techniques

In the third category, are classified computational intelligence-based techniques involving, amongst other concepts.

1.3.3.1. Artificial Neural Networks (ANNs)

In [33], an Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules, It consists to use an artificial neural network in order to estimate the output photovoltaic current and voltage under variable working conditions, The comparison between the estimated current and voltage with the ones measured gives useful information on the operating state of the considered photovoltaic module. The network inputs are solar irradiance and cell temperature, while its outputs are the PV current and voltage.

Another searchers cited in Ref [40], An intelligent method for fault diagnosis in photovoltaic array based on ANN is presented, this approach diagnosed degradation, shading and short circuit faults with requirement of three parameters in the input set: current of maximum power point (I_{mpp}), voltage of maximum power point (V_{mpp}) and the temperature of the PV modules. The three kinds of fault detected by ANN unit are

between Normal operation case, short circuit case, shading case and degradation case.

According to [46], the fault diagnosis of PV systems was important from point of view of detection, diagnosis and localization of various faults in PV generator based on ANN. In this approach, PV current, voltage and the number of peaks in the current-voltage (I-V) characteristic are computed based on a simulation model. The difference between the measured and simulated PV array output power is firstly calculated and compared with the threshold value (S), in the goal to detect faults occurrence. Then, the analysis of the main attributes, derived from (I-V) characteristic of each separate PV string, is elaborated to diagnose and localize the faults.

This method treated eight types of faults that have been precisely detected, diagnosed and localized in: module (short circuit in any cell or module or bypass diode, inversed bypass diode, shunted bypass diode and open circuit fault in module), connection fault with resistance between PV modules, partial shadow fault, shadow effect in a group of cells equipped by a faulted bypass diode and shadow effect in a group of modules connected by a resistance. This approach required two different algorithms:

- The first algorithm calculated the difference between the measured and the simulated PV array output power is compared with a threshold (Th) in order to detect the possible presence of a fault and it allows the discrimination of six faulty cases.

- While, the second algorithm is devoted to distinguishing between some faulty cases by the using of two types ANN classifiers (multi-layer perceptron (MLP) and radial basis function (RBF) classifiers requiring V_{mpp}, I_{mpp} and V_{oc} as input data.

An interesting method has been studied in Ref [47] under title Artificial Neural Network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. The types of neural network in this approach are Radial Basis Function (RBF) and Multilayer Perceptron Neural Network (MLP), which required two parameters solar irradiation and output power in its input. In this method, 10 various scenarios have been taken in consideration cited in table 1.1 knowing that the PV plant consists of 10 PV modules set-up in string topology.

Table 1.1. Various type of fault.

Cases number	Type of fault
1	Normal operation mode, where no faults were applied to the PV string
2	1 Fault applied to the system; 1 PV module disconnected from the PV string
3	2 Faults applied to the system;2 PV modules disconnected from the PV string
4	3 Faults applied to the system; 3 PV modules disconnected from the PV string
5	4 Faults applied to the system; 4 PV modules disconnected from the PV string
6	5 Faults applied to the system; 5 PV modules disconnected from the PV string
7	6 Faults applied to the system; 6 PV modules disconnected from the PV string
8	7 Faults applied to the system; 7 PV modules disconnected from the PV string
9	8 Faults applied to the system; 8 PV modules disconnected from the PV string
10	9 Faults applied to the system; 9 PV modules disconnected from the PV string

In Reference [50], two algorithms were proposed the first one focused on fault detection in PV generator using probabilistic neural network. The second diagnosed and located the frequent faults encountered in PV generator such as short circuit modules in string of PV array as well as the disconnected string in PV array using PNN. Both algorithms required in input set four parameters: cell temperature, solar irradiation, current and voltage of maximum power point. The types of faults considered in this approach are: short circuit diagnosis of three modules in PV string, short circuit diagnosis of ten modules in PV string and complete string disconnection diagnosis in PV array.

Other work demonstrates the effectiveness of Particle Swarm Optimization Back Propagation Neural Network (PSO-BP) which have been successfully used to detect and localise faults in PV array related to [40]. This approach required four parameters open-circuit voltage (V_{oc}), short-circuit current (I_{sc}), maximum power (P_{mpp}) and voltage at maximum power point (V_{mpp}) are extracted from the output curve of the PV array as identification parameters for the fault diagnosis system. This method diagnosed six different faults in PV generator citing PV temperature fault, partial shade fault, aging cells, the combination of temperature and shade, the combination of temperature fault and aging cells, and the combination of shading and aging cells.

Probabilistic Neural Network algorithm has been combined to Fuzzy C means and Gaussian kernel algorithms into a novel PV array fault diagnosis method demonstrated in [51]. Real irradiance and temperature data were input to the model to simulate PV array output characteristics under different conditions. The PNN required four input data I_{mpp}, V_{mpp}, V_{oc} and I_{sc} to diagnose: short circuit fault of two, four and six PV modules in PV array, open circuit fault of one string and two strings in PV array, shading fault of two, four and six modules and finally abnormal aging fault with 2, 4 and 6Ω.

1.3.3.2. Neuro Fuzzy logic

A hybrid Neuro-Fuzzy approach of fault detection and diagnosis has been proposed in [43]. This approach needs three important steps: firstly a Neuro-Fuzzy model of PV modules has been developed, secondly the six attributes (Isc, Voc, Impp, Vmpp, S1 which represents the incremental derivative ratio calculated considering the following relevant points in the I-V curve: short current point (0; Isc) and maximum power point (Vmpp ; Impp); and S2 which represents the incremental derivative ratio considering the two points (Vmpp ; Impp) and open circuit (Voc ; 0).) using the I-V characteristic have been analyzed and extracted, and at last the application of Norm-test has been used to detect and diagnose faults. According to this work, three Neuro-Fuzzy models (blocks) have been developed to model the system under ideal, healthy and faulty conditions. The ideal condition block defines the case when all the inputs and outputs are considered perfectly constants and noiseless. In this work, five operating cases: normal operating case, diode short-circuit operating case, lower earth fault operating case, upper earth fault operating case and partial shading operating case .

An interesting comparative method between Mamdani, Sugeno fuzzy logic and radial basis function RBF ANN network for PV fault detection has been demonstrated in reference [44]. The fault detection algorithm can detect and locate accurately eleven different types of faults occurring in PV array that is illustrated in table 1.2.

Table 1.2. Different type of faults occurring in the examined PV plant [42].

Type of Fault	**Symbol**
Normal Operation and PS effects the PV system	F1
One faulty PV module	F2
Two faulty PV module	F3
Three faulty PV module	F4
Four faulty PV module	F5
One faulty PV module and PS effects the PV system	F6
Two faulty PV module and PS effects the PV system	F7
Three faulty PV module and PS effects the PV system	F8
Four faulty PV module and PS effects the PV system	F9
Faulty PV String	F10
Faulty MPPT unit	F11

In the other side, a well-functioning procedure has been proposed in ref [45] of fault detection based on curves modelling and fuzzy classification system. In this approach LabVIEW software has been used to simulate the meteorological conditions (T and G), the voltage ratio VR, and the power ratio PR. This method used a third polynomial function to compute two detection limits (high and low detection limits) of VR and PR ratios. After that, these limits are compared with real measured data of an actual PV generator. Samples that lie out of these limits are then introduced to a fuzzy-logic based classification system to diagnose the fault type.

Noting that, this method could successfully detect the frequent faults such as:

- The presence of partial shading (PS) within the PV generator.
- One short-circuited PV module and PS.
- Two short-circuited PV modules and PS.
- (Q-1) short-circuited PV modules and PS, where Q denotes the total number of PV modules.

The principal weak point of this method depends on power and voltage ratios. Consequently, its efficiency depends highly on the robustness of the instrumentation components.

1.3.3.3. Decision Tree

In [52], an attractive method of AI has been proposed to detect and diagnose fault in GCPV systems based on Decision Tree under title “Fault detection and diagnosis based on decision tree algorithm for grid connected PV system”. This method diagnoses the most frequent faults encountered in PV installations such as: open circuit fault, short circuit fault and line-line fault and requires three attributes temperature ambient, irradiation and power ratio calculated from measured and estimated power collected from Sandia model. Two targets are proposed in this approach: the first target is between two positions either healthy or faulty state for detection. While the second target is to diagnose four classes between healthy and the three classes cited above.

1.4. Conclusion

This chapter has exposed the various fault between temporary and permanent internal or external encountered in PV systems which have an important impact on the reliability and efficiency in output energy of PV plant. And finally, the three diverse categories of fault detection and diagnosis in PV systems according to the literature, citing: Model based diagnosis technique, signal processing based fault diagnosis methods and artificial intelligence in fault diagnosis techniques.

According to the limitations imposed by these categories, our opinion has been oriented towards the classification of faults in PV array. In this aim, modeling the behavior of PV generator in healthy and faulty cases is essential which represents the objective of the next chapter where Simscape MATLAB/Simulink is present.

CHAPTER 2 MODELLING AND SIMULATION OF PHOTOVOLTAIC GENERATOR

2.1. Introduction

The aim of this work is to develop an intelligent tool capable to diagnosis a grid connected PV plant. For this, the first grid connected PV plant in Algeria has been considered. In this chapter, a detailed presentation of this PV plant is given, its technical specification and the experimental data.

The development of the proposed diagnosis method went through several stages, the first one is the methodology and the strategy which is very important, as it is considered as the starting phase, where a creation of a significant database under different cases between healthy and faulty in a PV generator has been elaborated. The proposed plan is divided under five main process:1- Presentation of experimental PV plant and physical model of PV cell, 2- Simulation and experimental validation of PV array using Simscape MATLAB/ simulink as well as the elaboration of pertinent database including healthy and faulty operations such as: one PV module short circuit, two PV modules short circuit, four PV modules short circuit and faulty string with the same working conditions, 3- An analytical study will be presented faulty behavior based on the deviation between the desired output and the experimental real output under the same working conditions.

2.2. Methodology

Figure 2.1 shows the principal steps procedure that is considered as the basic step in this research. This part needs real PV generator and simulated PV generator with the same PV modules number and the same characteristics of both types. The aim of this step is firstly to build a database set with different cases between healthy and faulty models and secondly to proceed to the fault detection procedure by calculating residual information which require a comparative study between real measured data collected from PV station and simulated data collected from simulated model. It is an optimal part that requests the calculation of different criteria.

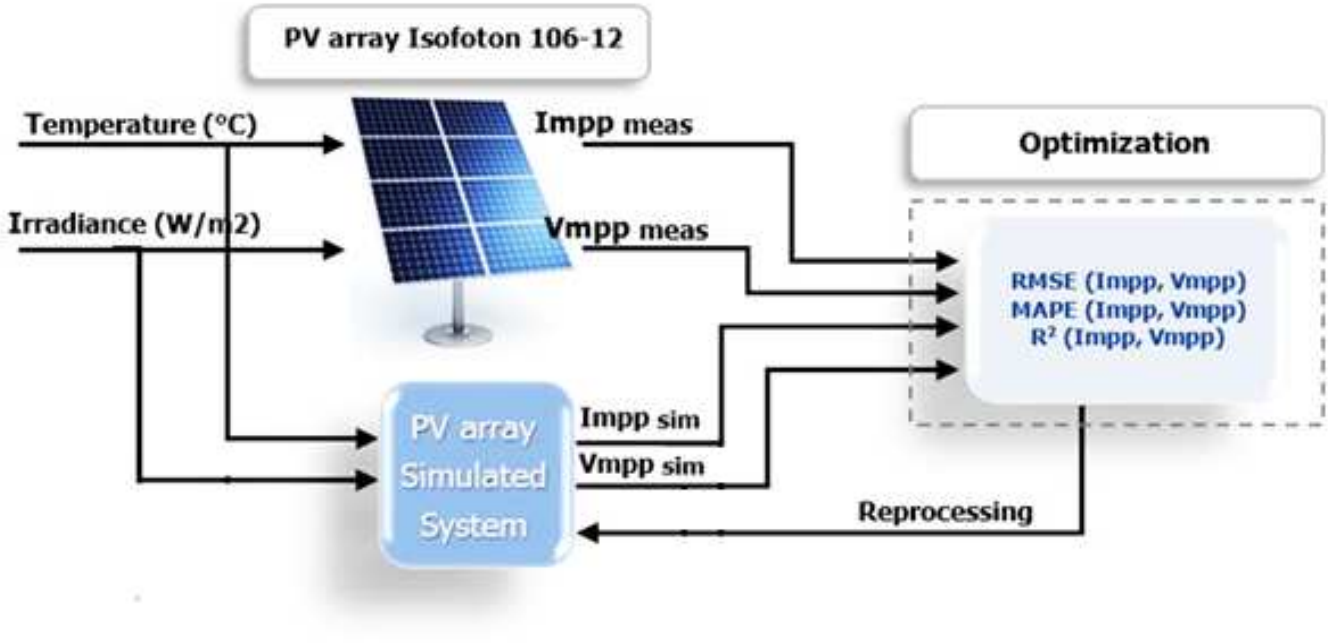


Figure 2.1. Scheme of PV model parameters identification.

2.3. Presentation of experimental PV plant

The current study employs an experimental setup situated in the capital Algiers of Algeria. The installation of grid connected PV system is part of the Algerian-Spanish cooperation placed on the roof and is composed by 90 monocrystalline PV modules divided in three sub-array each one assembles 30 PV modules illustrated in figure 2.2, where each sub-array is linked to an inverter. The global PV generator provides output power equal to 9.54 kiloWatt-peak injected in a grid of 220V, which implies that the output power of each sub-array is equivalent to 3.18kWp. The connection of PV modules in each sub-array is in 2 parallel strings composed by 15 PV modules arranged is series. The brand of PV modules in this installation is Isofoton106-12 its characteristics is mentioned in table 3.1, while for the inverter is IG30 Fronius its - Max Input = 3.6KW and Max-Output=2.65KW. The PV plant requires two major parameters at its input setup, PV module temperature that is measured grace to K-type thermocouple as well as the inclined and horizontal solar irradiation measured by Kipp & Zonen CM11 thermoelectric pyranometer. Data acquisition can be configured using Agilent 34970. Figure 2.3 represents the major's components of small grid connected PV systems encounter in this study.

Figure 2.2. Roof grid connected PV plant in Algiers, Algeria.

Table 2.1. Electrical properties of the Isofoton 106-12 PV module (STD: solar irradiation=1000W/m2, cell temperature= 25°C) [17, 37, 77].

Solar Panel electrical characteristics	Value
Peak power	106 W
Short circuit current (Isc)	6.54 A
Open circuit voltage (Voc)	21.6 V
Voltage at Maximum Power Point (Vmpp)	17.4 V
Current at Maximum Power Point (Impp)	6.10 A
Number of cells connected in Series	36
Number of cells connected in Parallel	2
Cell Short circuit current	3.27 A
Cell Open circuit Voltage	0.6 V

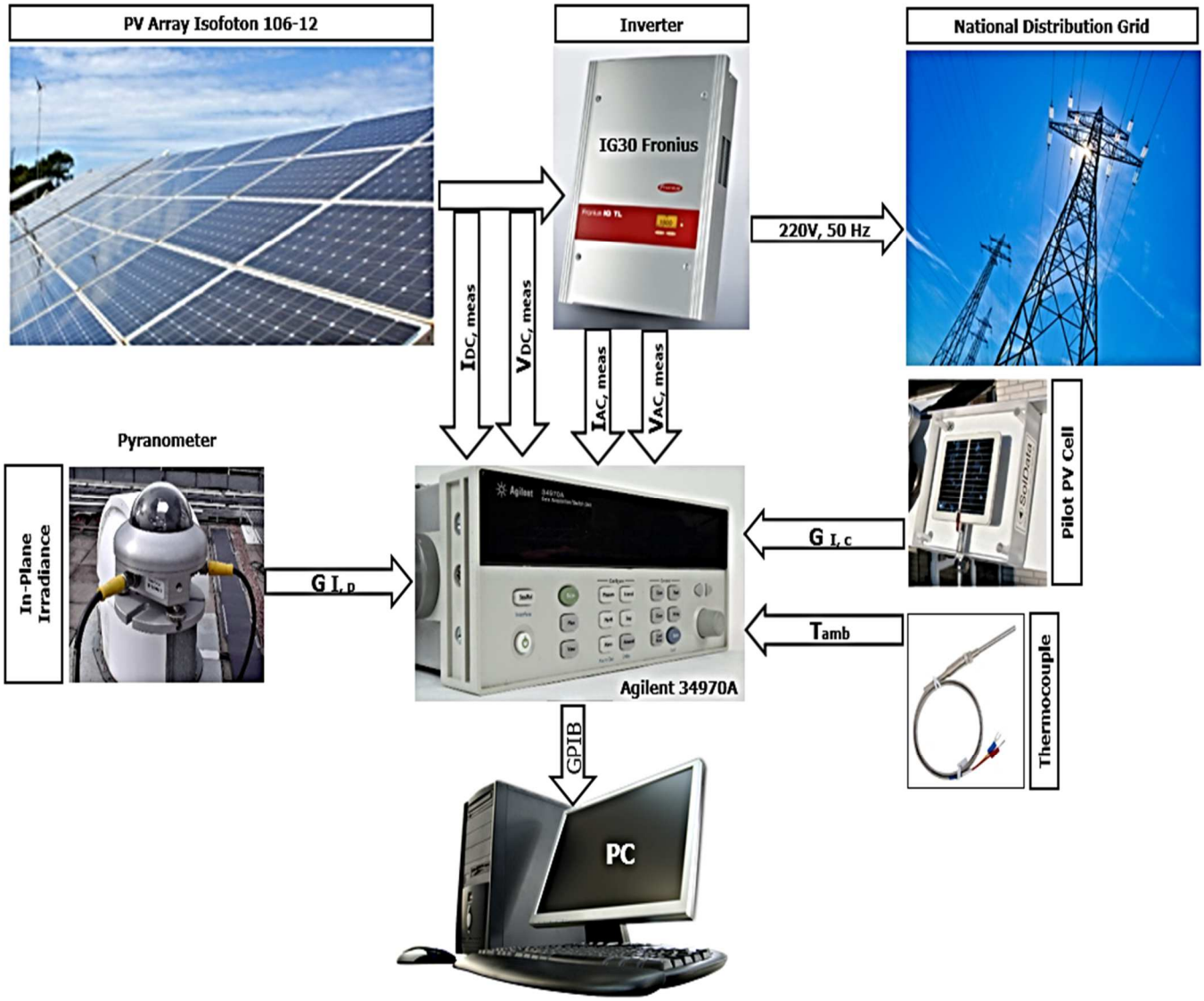
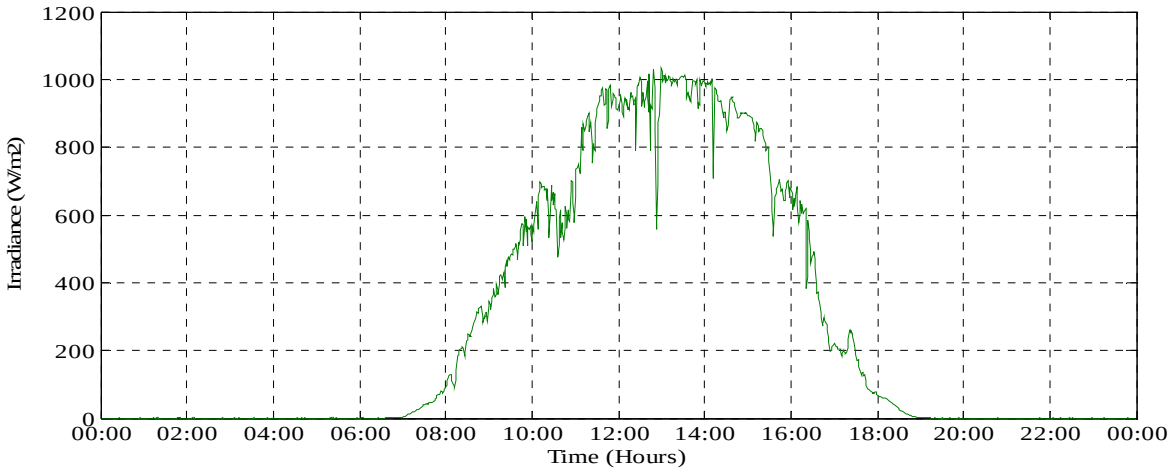
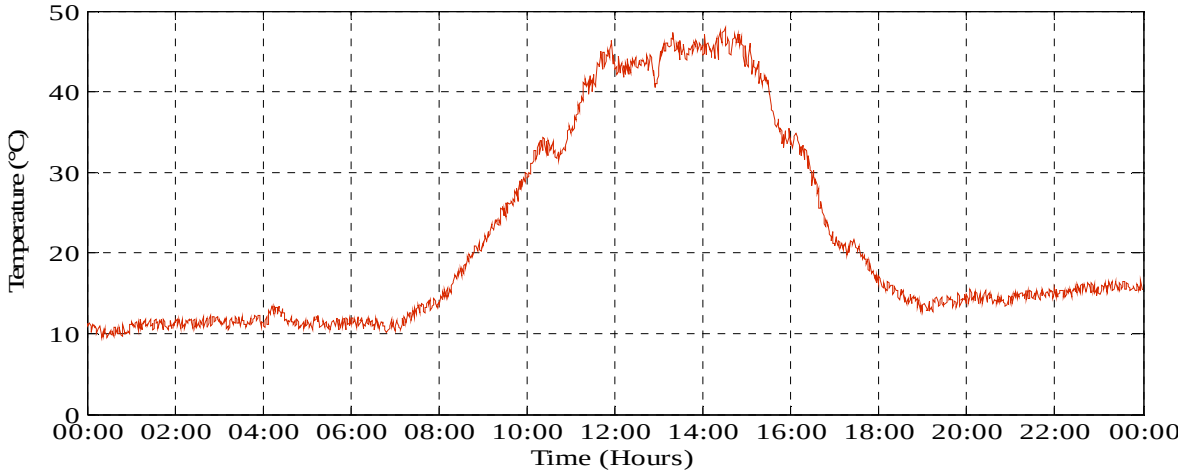


Figure 2.3. Experimental setup.

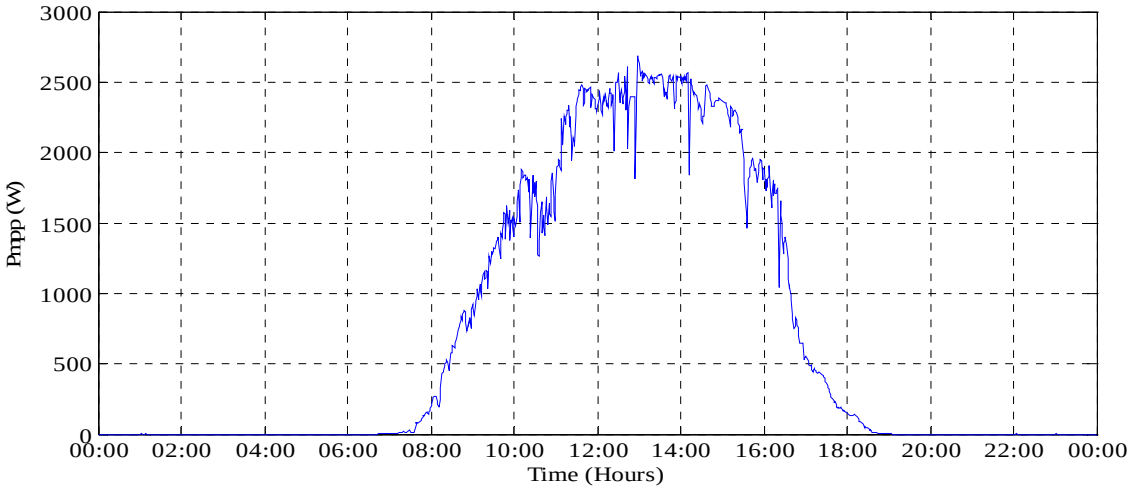
Figure 2.4 illustrates the output power of the setup and the corresponding weather conditions (solar irradiation and ambient temperature) during one day (24 hours) collected from the experimental setup.



(a)



(b)



(c)

Figure 2.4. Daily profile of onsite measured (a) Irradiance, (b) Temperature, (c) Output power.

2.4. PV array modelling:

Model based fault diagnosis need accurate models to perform fault diagnosis. Here, the mathematical model outputs of the PV array are needed to be compared with the outputs of the PV array under study. It is important to note that model based diagnosis is highly dependent on the accuracy and on the complexity of the used model. PV array modelling has been performed in three steps:

2.4.1. Modeling of PV cell

The PV cells, also called the solar cells is considered as basis device in PV installation that convert sunlight energy in electrical energy via semi-conductor based on P-N junction by the PV effect. These PV cell or PV modules can contain one diode or two diodes, the most popular are one diode model.

In this study, one diode model has been used illustrated in figure 2.5 that represents the most popular physical model. Mainly, the model was developed for only one cell. Its generalization to all module involved that all cells are considered strictly identical. This model contains five important parameters (I_0, I_{ph}, R_s, R_{sh} and n) giving an output current via an output voltage demonstrated in the following equation [62, 63]:

$$I = Iph - I_0 \left(exp \frac{q(V+R_s I)}{nkT_c} - 1 \right) - \frac{V+IR_s}{R_{sh}} \qquad (2.1)$$

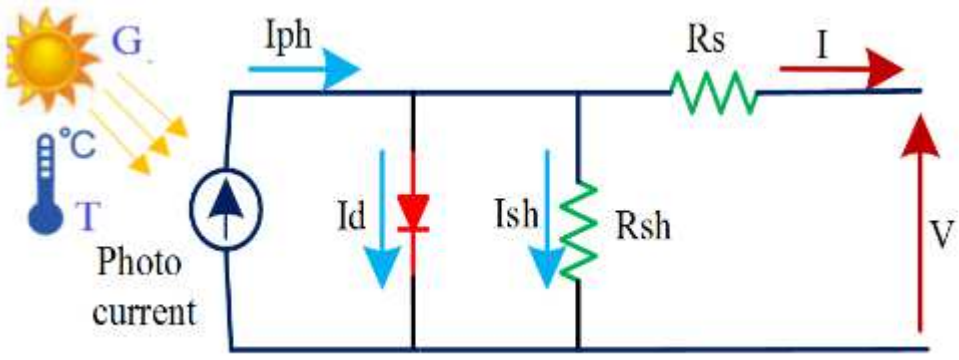


Figure 2.5. Equivalent circuit of solar cell [41].

Where :

I and V : are respectively the output current and voltage of PV module.

I_0 : is the diode saturation current.

I_{ph} : is the photo generated current.

R_s : serie resistance.

R_{sh} : shunt resistance (parallèle)

T_c : PV cell or module temperature.

k : Boltzman constant (1.3806503 x 10-23 J/°K)

q :electron charge (1.6 0217646 x 10-19 C)

n : diode ideality factor.

This model is called 5 parameters one diode model. The five parameters are I_0, I_{ph}, R_s, R_{sh} and n. Modelling the PV module/sub array using this model is made by finding the best combination of these 5 parameters so the output of the model can give the best fit of the experimental data.

2.4.2. PV module / array modelling and I-V characteristics:

Figure 2.6 presents a nonlinear I-V characteristic of PV module. These characteristics are dependent on solar irradiation level and ambient temperature as illustrated by Figure 2.6. These two weather variables are called working conditions that are the inputs of the PV module/array model while the I-V curve is its output. The I-V curve is characterized by three main points; the short circuit current I_{sc}, the open circuit voltage V_{oc} and the maximum power point (MPP) at which the power is at its maximum. The current and the voltage at the MPP are called V_{mpp} and I_{mpp} respectively. The knee of the I-V curve symbolizes the maximum power point (P_{mpp}) of the PV module that is generated in standard condition. Knowing that the quality of module technology and manufacture owns serious influences on efficiency of electrical output power.

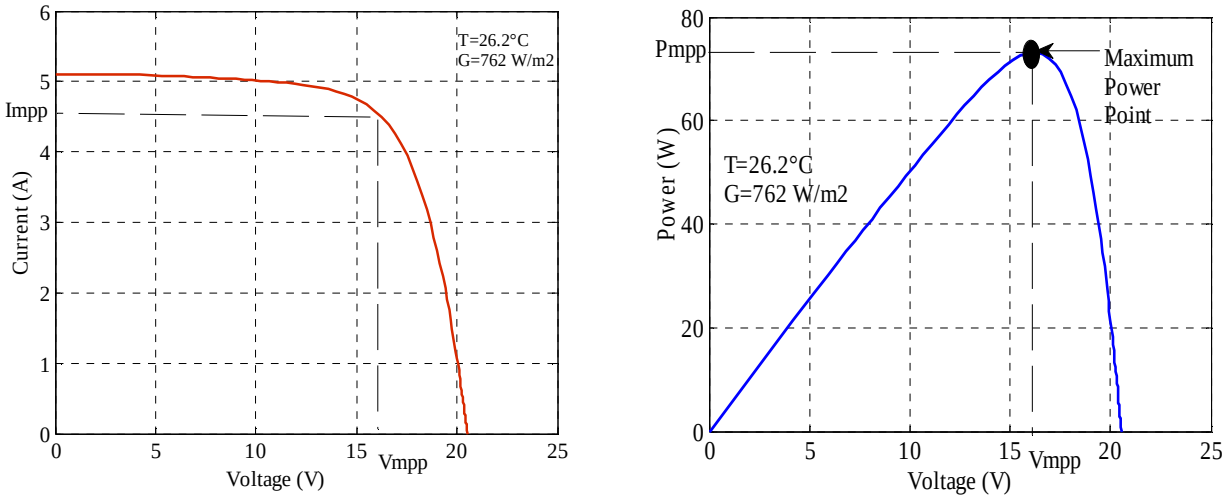
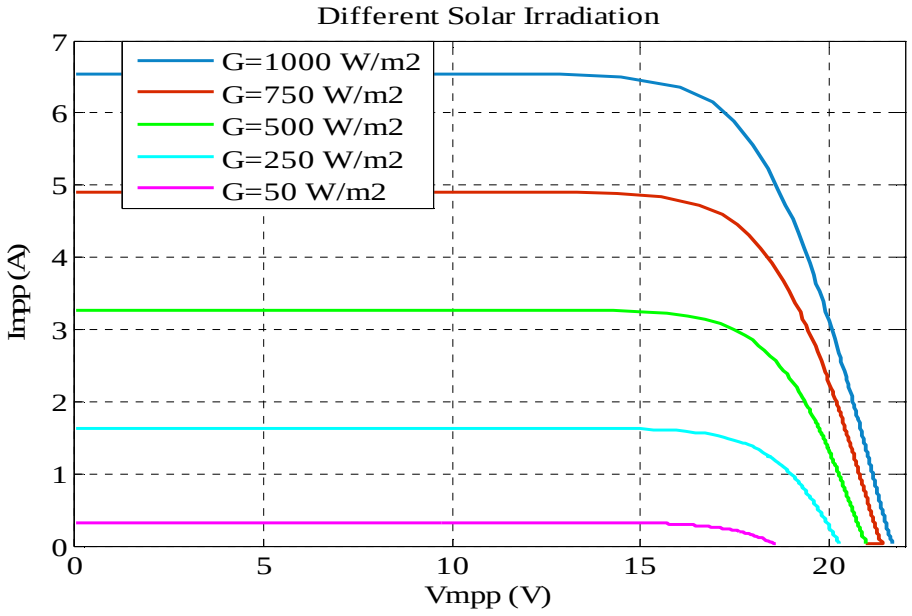
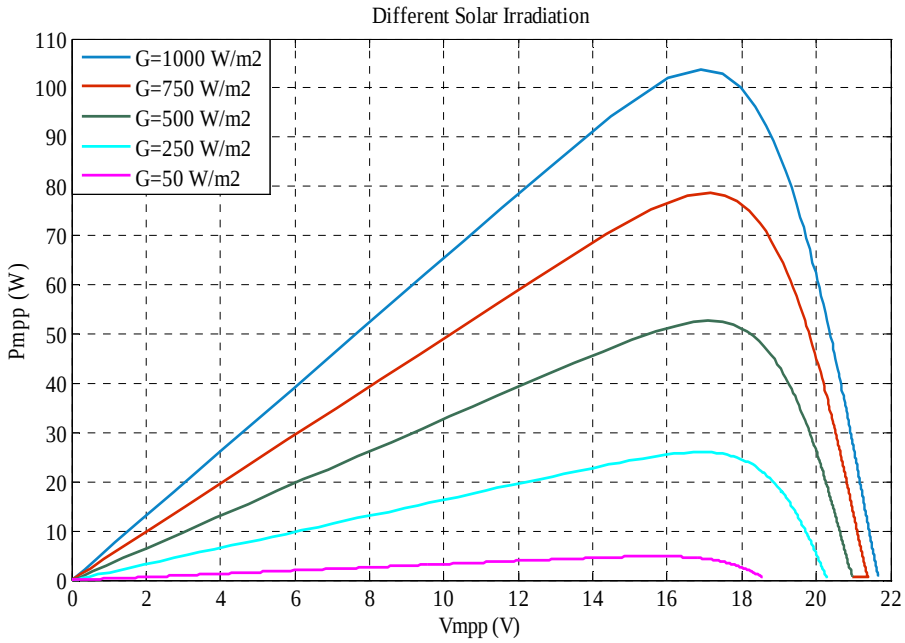


Figure 2.6. I-V and P-V curves.

Figure 2.7 and Figure 2.8 represent the I-V and P-V curve influences by the weather conditions, the two first (a) and (b) of figure 2.7 represent the different change of solar irradiation with constant cell temperature that has a remarkable effect on short-circuit current therefore in the output power, but negligible effect on the open circuit voltage. The last two (a) and (b) of figure 2.8 represent the different variations of cell temperature and constant solar irradiation. It is clear that the change in temperature has a strong effect on the open-circuit voltage also the output power of the PV cell, contrary to short circuit current, which has negligible effect on the output system.

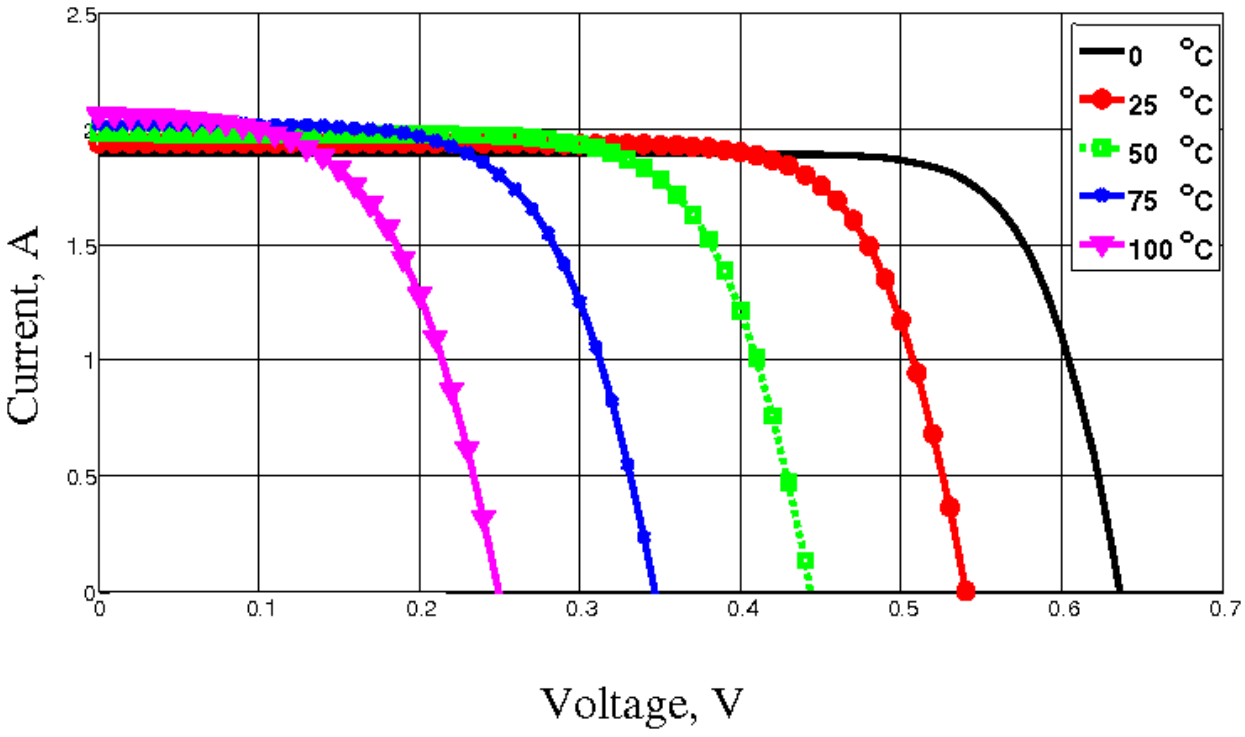


(a)

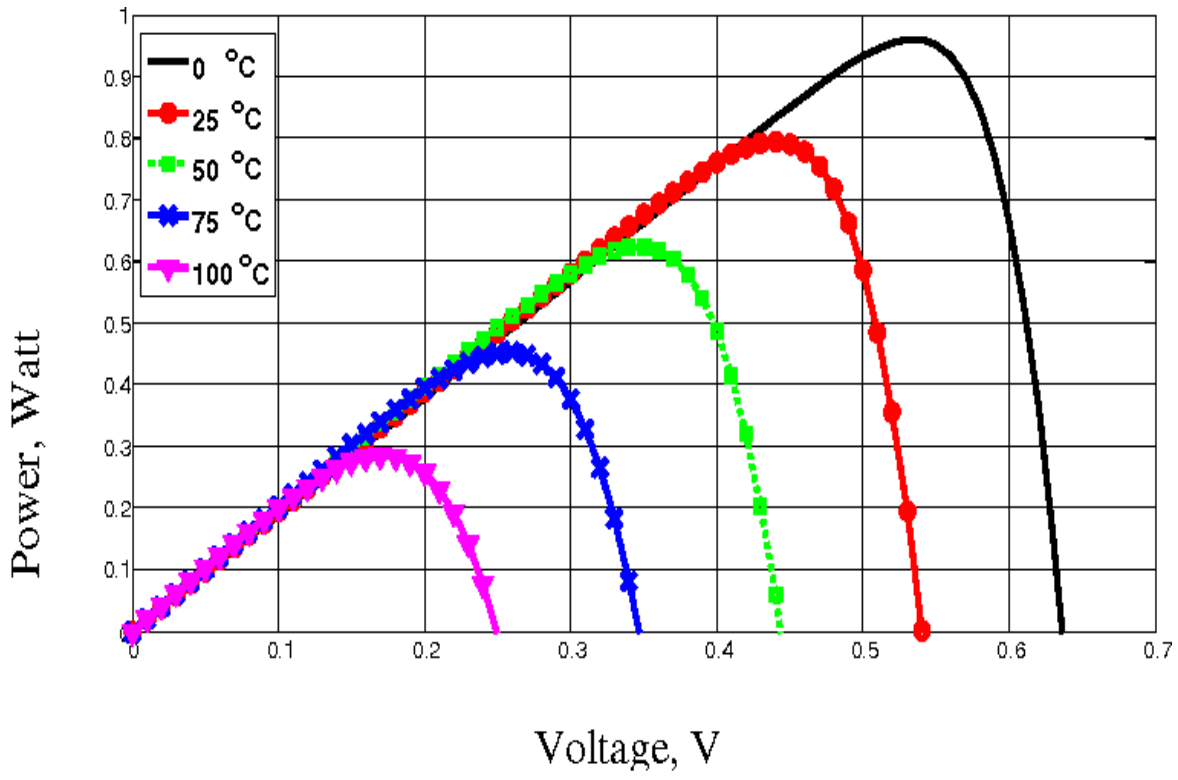


(b)

Figure 2.7. Effect of Solar Irradiation (a) I-V and (b) P-V curves (the following).



(a)



(b)

Figure 2.8. Effect of Cell Temperature (a) I-V and (b) P-V curves.

2.5. PV array model simulation:

PV array modeling has been carried out using Matlab software. Two simulation strategies are possible. The first is the simulation of equivalent circuit model by functional equations using the script language of Matlab. The second is the simulation of the equivalent circuit model blocks using Simscape. The advantage of Simscape is that it enables to rapidly create models of physical systems within the Simulink environment. With Simscape, it is possible to build physical component models based on physical connections that directly integrate with block diagrams and other modeling paradigms [61]. For this work, the second option has been chosen because it allows to simulate the targeted faults as the short circuit and the disconnection faults. Figure 2.9 illustrates the interconnected of 16 PV modules based sub array linked in two strings, each string is composed by 8 PV modules (2 x 8). Each PV module receives two climatic data: cell temperature and solar irradiance.

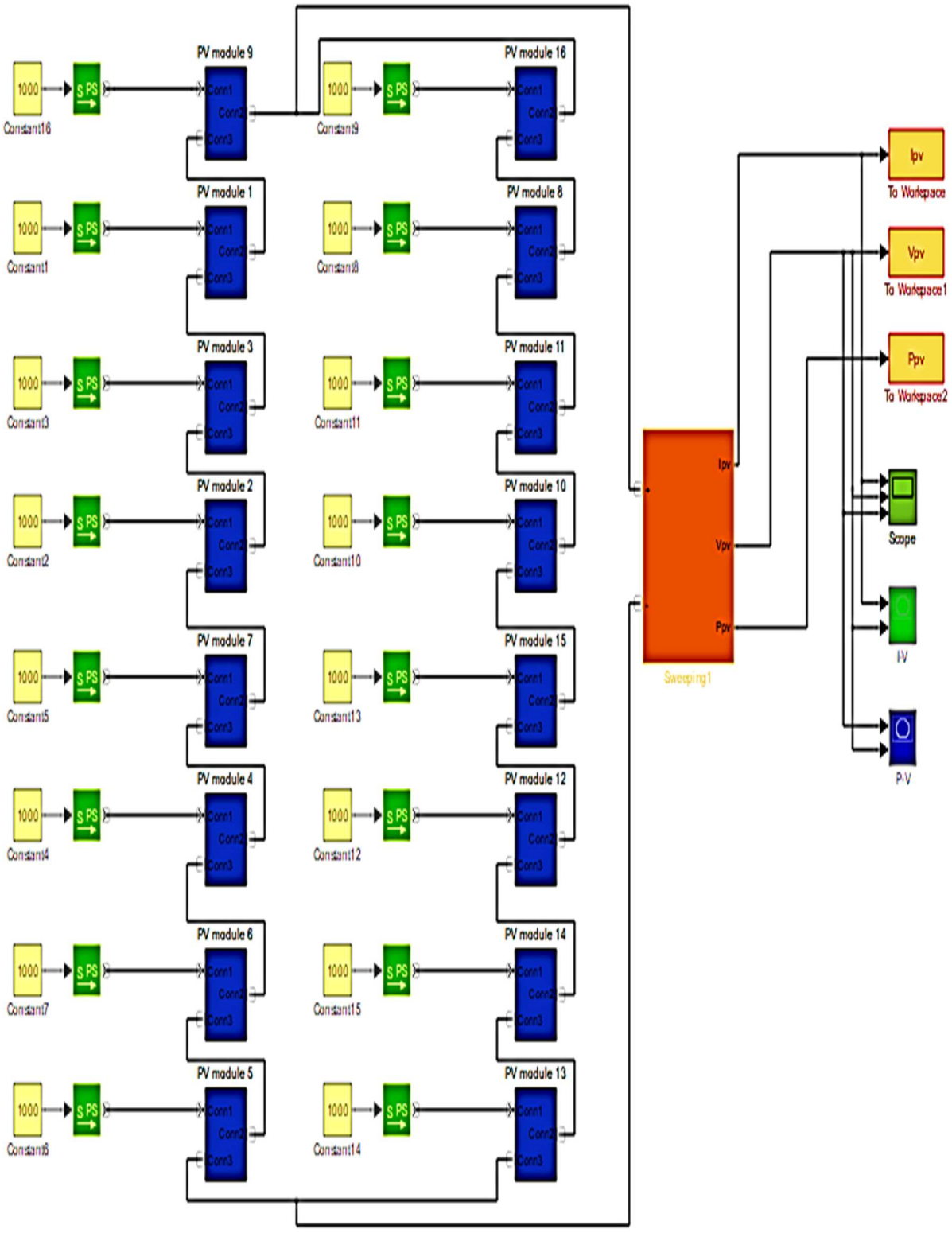
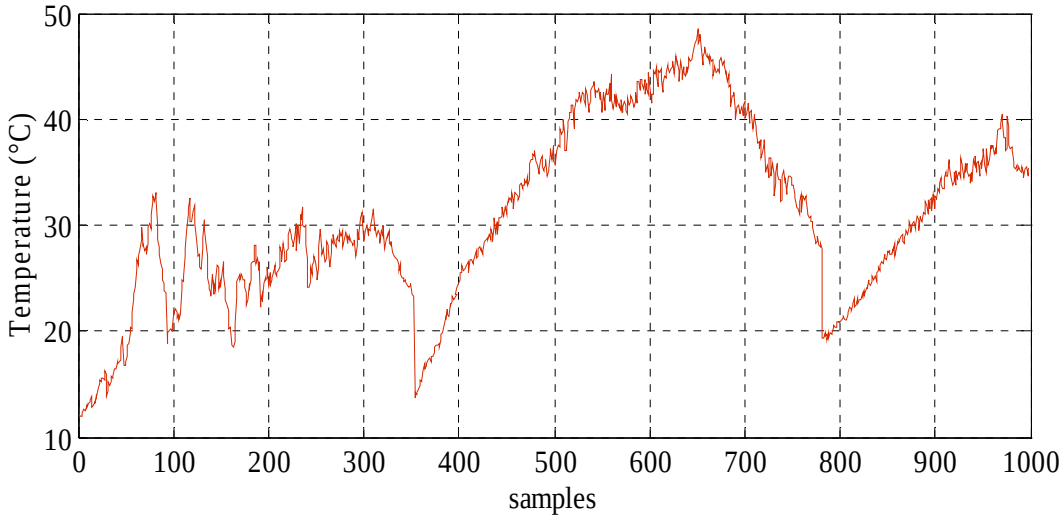


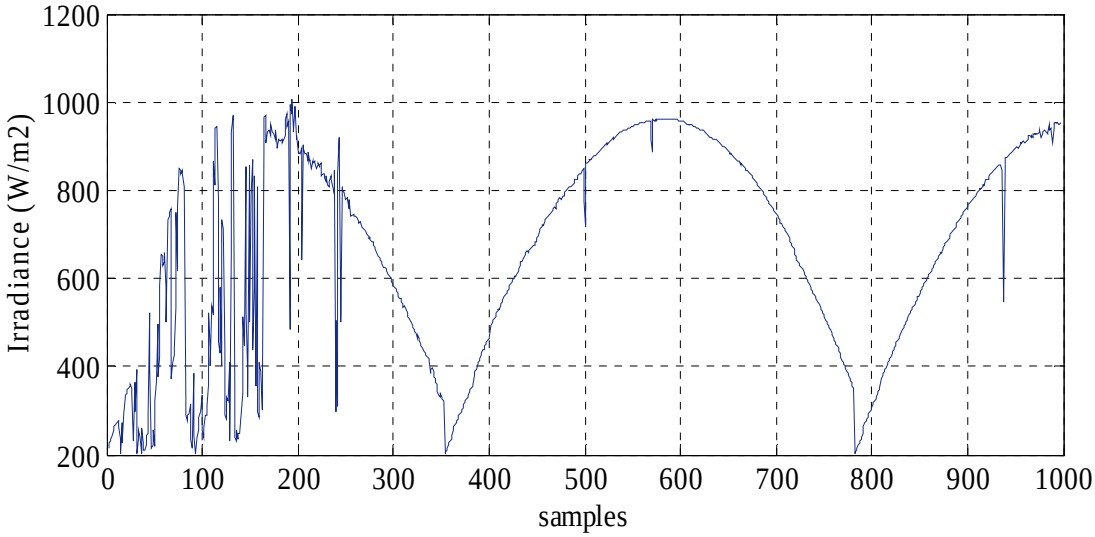
Figure 2.9. Block diagram of the studied PV sub-array in Simscape /MATLAB.

2.5.1. PV array data validation:

The validation of the obtained model has been done using experimental data collected in March month from the experimental setup situated in the capital Algiers of Algeria. Note that the validation data are the inputs of the models which are the working condition (the solar irradiation and the ambient temperature), and the output data are the current and the voltage at the MPP as illustrated by Figure 2.10. Knowing that the experimental setup received one sample each minute and the efficiency of the studied model required data above 200 W/m^2.

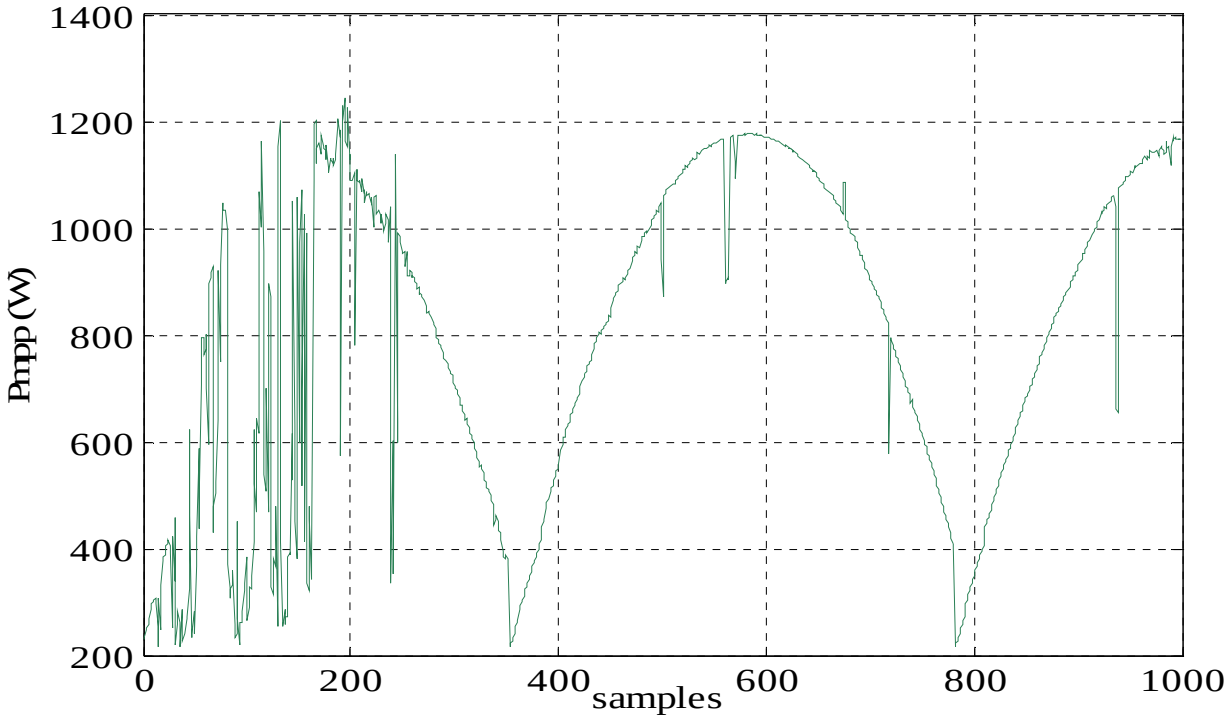


a) The temperature

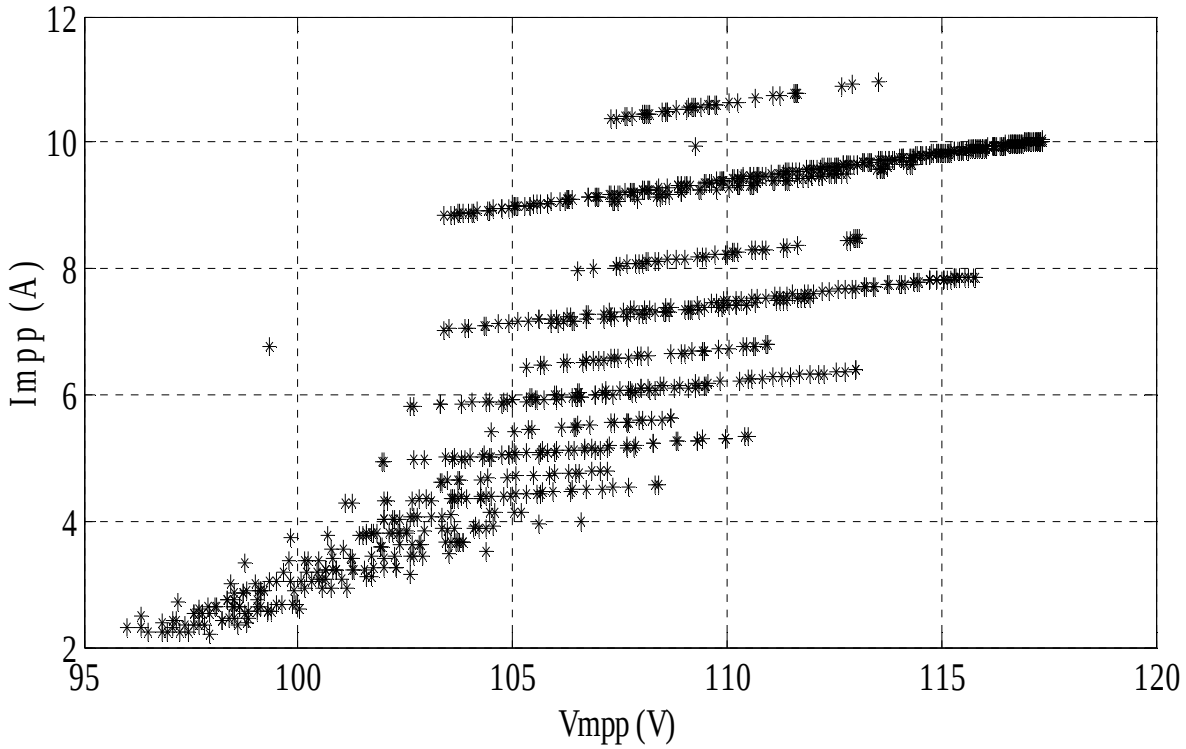


b) The irradiance.

Figure 2.10. Experimental validation data.



c) The Power at MPP.



d) The current and voltage at the MPP

Figure 2.10. Experimental validation data (the following).

2.5.2. Healthy real sub array versus healthy simulated sub array

In order to validate and ensure the accuracy and precision of the simulated generator, the study take in consideration three evaluation criterions to compare the real measured data with simulated data obtained from the present simulated model, where each criterion described as follow:

- Root Mean Square Error (RMSE) describes the difference between the real experimental data of the PV system and simulated data obtained from Simulink/ PV generator related to the used number of samples [16, 77].
- Mean absolute percentage error (MAPE) describes the difference between the real experimental data of the PV system and the simulated data obtained from Simulink/ PV generator then the result is divided by the real experimental data of the PV system, after its modulus is divided on the used number of samples. The advantage of this tool is its easy understanding gave in percentage [17, 77].
- The coefficient of determination is a statistic that will give some information about the goodness of fit of a model [60], which explains by linear regression, its range is 0 to 1, if this coefficient is 1 or very close it means that the result is ideal. The result of this tool is sum of square of difference between the real experimental data of the PV system and the simulated data obtained from Simulink/ PV generator, the result is divided on the sum of square of the real experimental data of the PV system, after that the new result will be put in negative form plus 1.

The equations of these criterions are represented respectively as following [16, 50, 51, 77].

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \widehat{y_n})} \quad (2.2)$$

$$MAPE = \frac{1}{n}\sum_{n=1}^{N}\frac{|y_n - \widehat{y_n}|}{|y_n|}x100\% \quad (2.3)$$

$$R^2 = 1 - \left(\frac{\sum_{n=1}^{N}(y_n - \widehat{y_n})^2}{\sum_{n=1}^{N}(y_n)^2}\right) \quad (2.4)$$

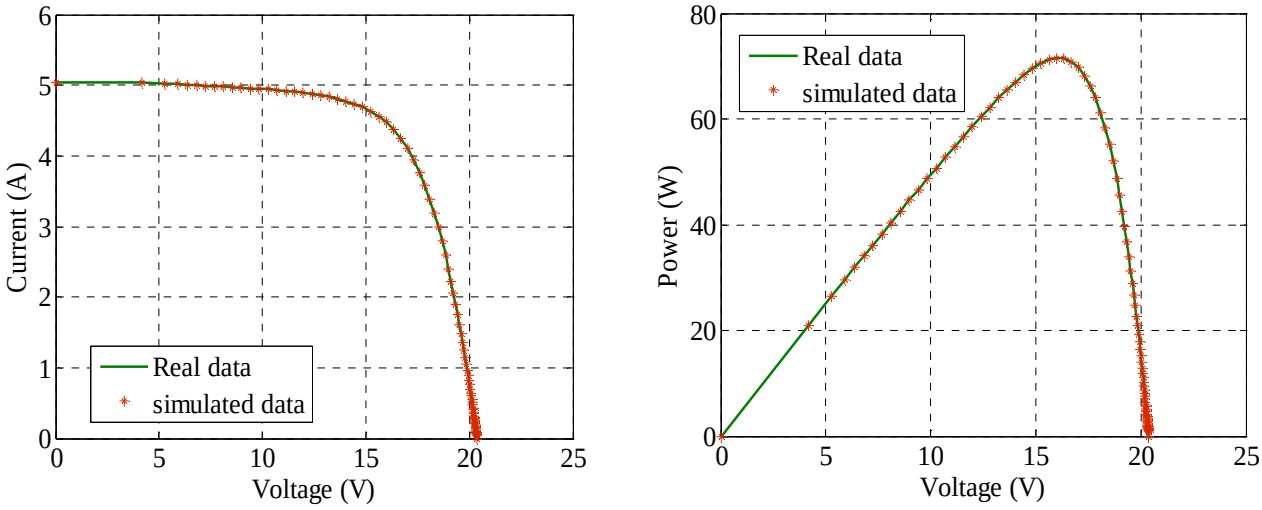
Where:

y_n: the nth measured data.

$\hat{y}_n$: the nth simulated data.

N: the size of the database (the number of the validation's samples).

Consequently, the effectiveness of the IFD Approach relies at once on its PV module's modelling precision and faults classification accuracy. For the first criterion, the obtained results show good agreement between the measured data and the model generated data as illustrated by Figure 2.11.



a) I-V curve b) P-V curve

Figure 2.11 Measured and simulated PV module output data.

The corresponding error metrics are summarized in Table 2.2. The correlation coefficient between the real and simulated curves is 99 % for both voltage and current, while the Root mean square error is around 5 % and Mean Absolute Percentage Error is between 2 and 5 for both voltage and current of maximum power point.

Table 2.2. Model performance metrics values.

	RMSE (%)	MAPE (%)	R^2 (%)
Impp	04.33	02.06	99.99
Vmpp	06.06	05.23	99.26

2.5.3. Current validation of maximum power point.

Figure 2.12 demonstrates the measured current against the simulated current of maximum power point on a number of 1000 samples. The result show a high efficiency of simulated model, where its projection on real measured model shows perfect proof reveals in Figure 2.13 that represents zoom in part of Figure 2.12. The gap by absolute error between measured and simulated current do not exceed 0.02 A for current at MPP as shown in Figure 2.14.

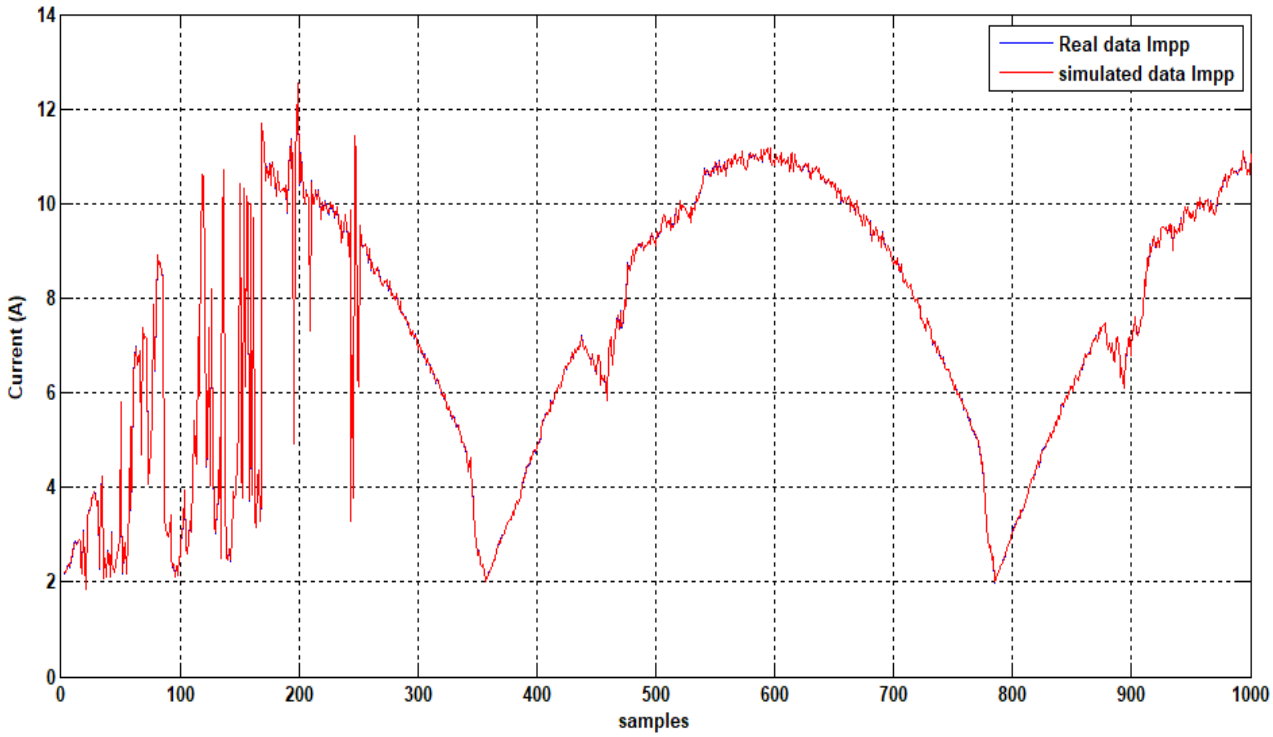


Figure 2.12. Current of real measured data against simulated data of PV generator.

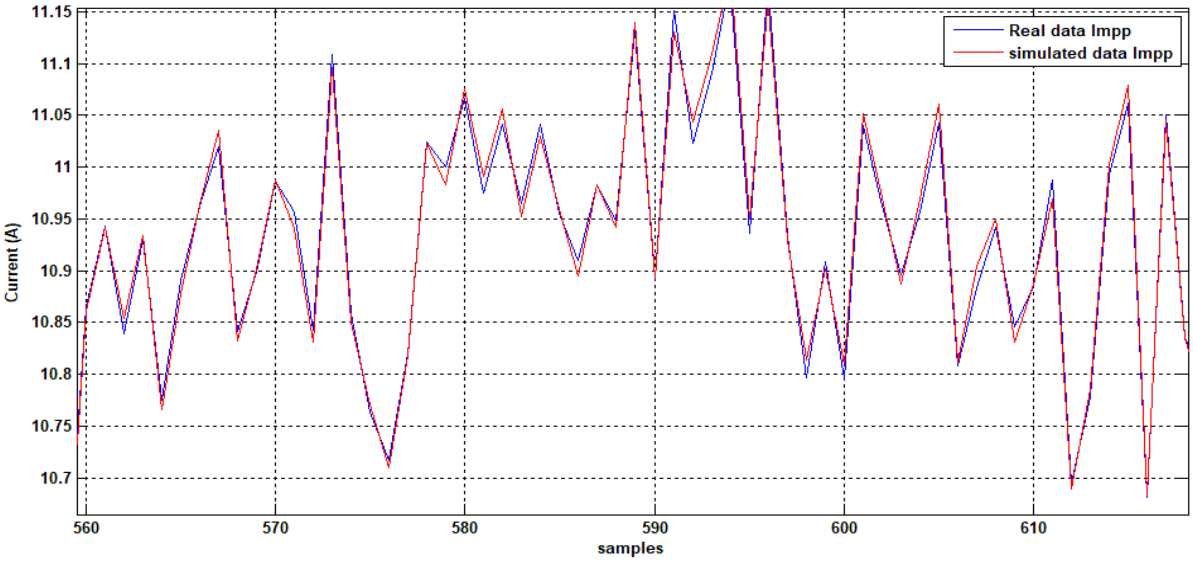


Figure 2.13. Part of figures 2.12 zoomed.

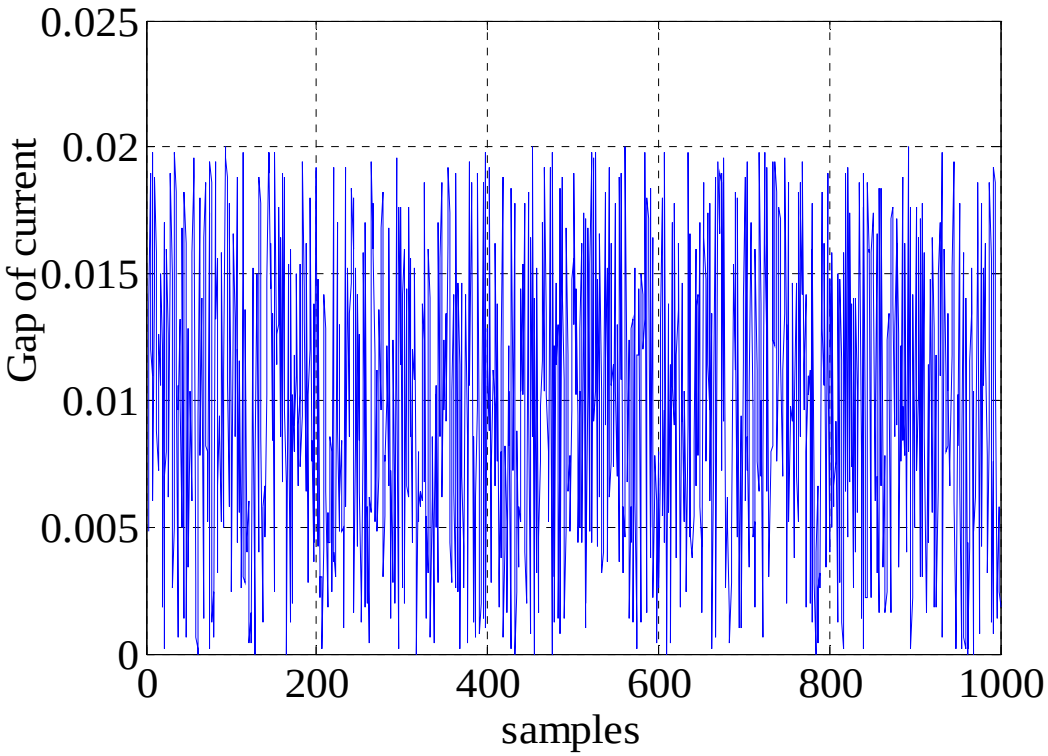


Figure 2.14. Absolute error between measured and simulated current at MPP.

The PV generator modeling accuracy is the sole guarantee for MPP identification and thus of the faults detection. The identification results are traced as simulated data versus measured data of the MPP current (Class1I) shown in Figure 2.15.

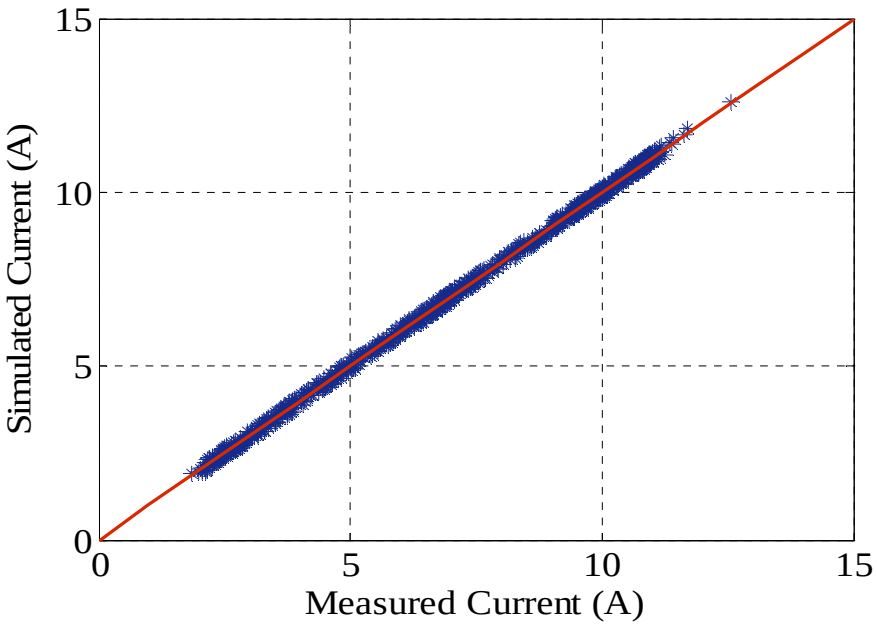


Figure 2.15. Identification of the current of maximum power point in normal operation conditions.

2.5.4. Voltage validation of maximum power point.

As represented in Figure 2.16, the obtained results for the second electrical parameter show a high agreement between the voltage of real measured data and the simulated data versus 1000 samples, where the Figure 2.17 illustrates a good precision zoomed of Figure 2.16.

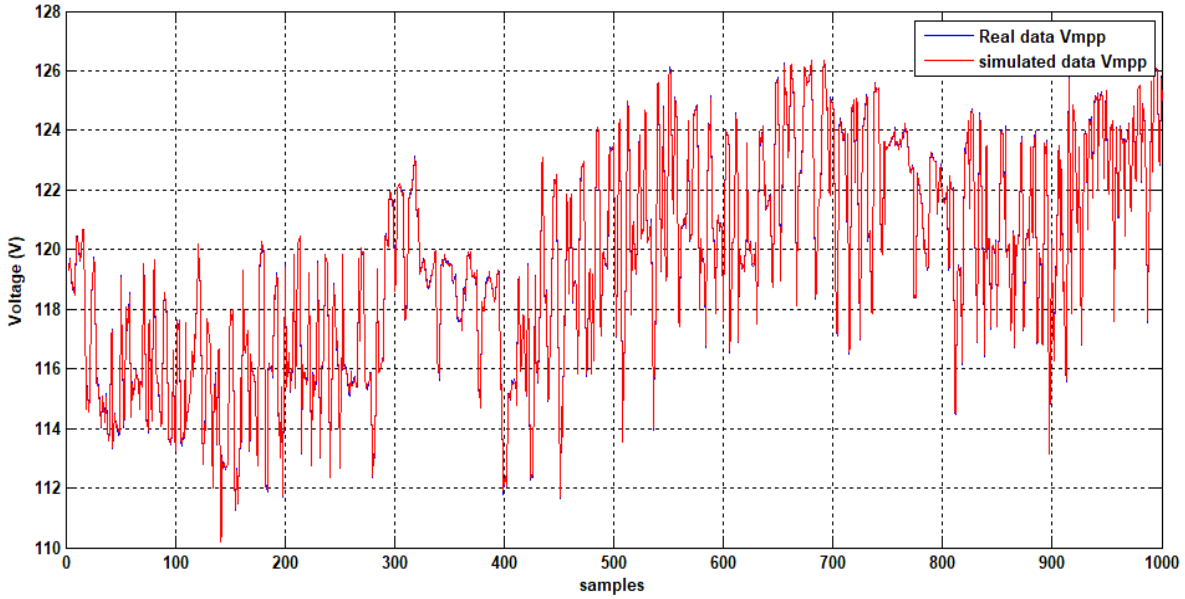


Figure 2.16. Voltage Measured real data against simulated data of PV generator.

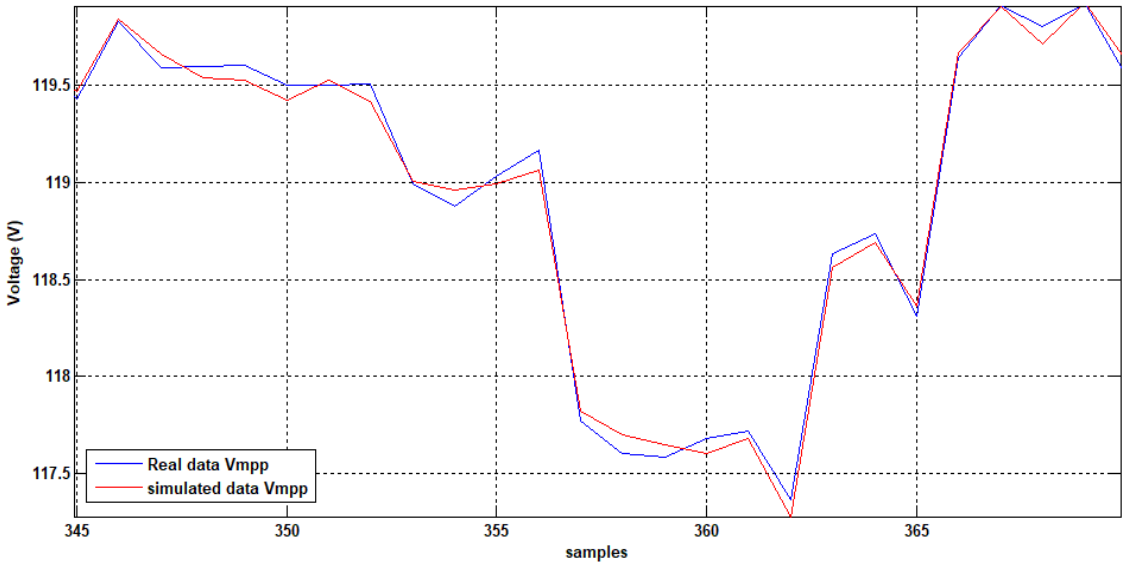


Figure 2.17. Part of figure 2.21 zoomed.

The gap by absolute error between measured and simulated data do not surpass 0.11V for voltage at MPP shown in Figure 2.18. The identification results is traced as simulated data versus measured data of the MPP voltage (Class1V) shown in Figure 2.19.

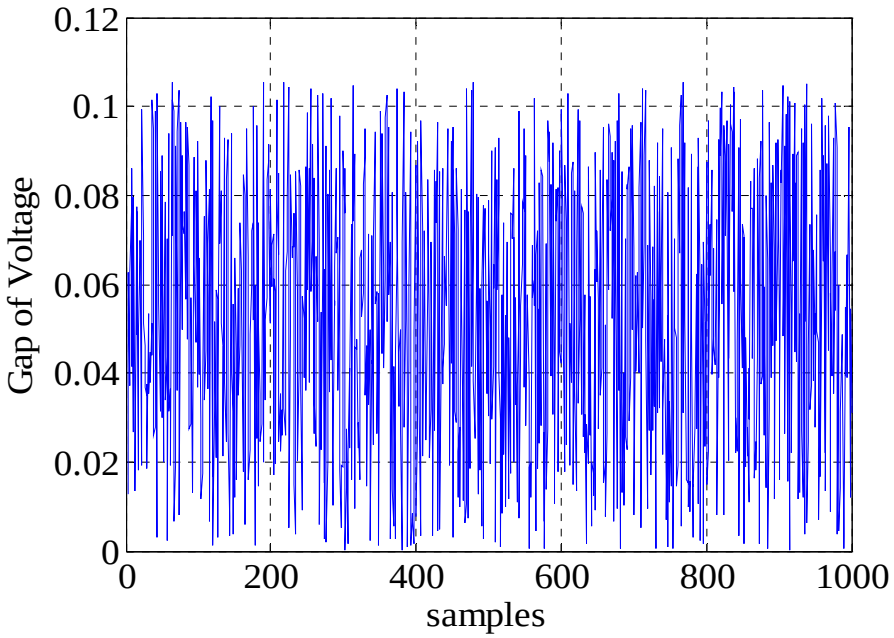


Figure 2.18. Absolute error between measured and simulated voltage at MPP.

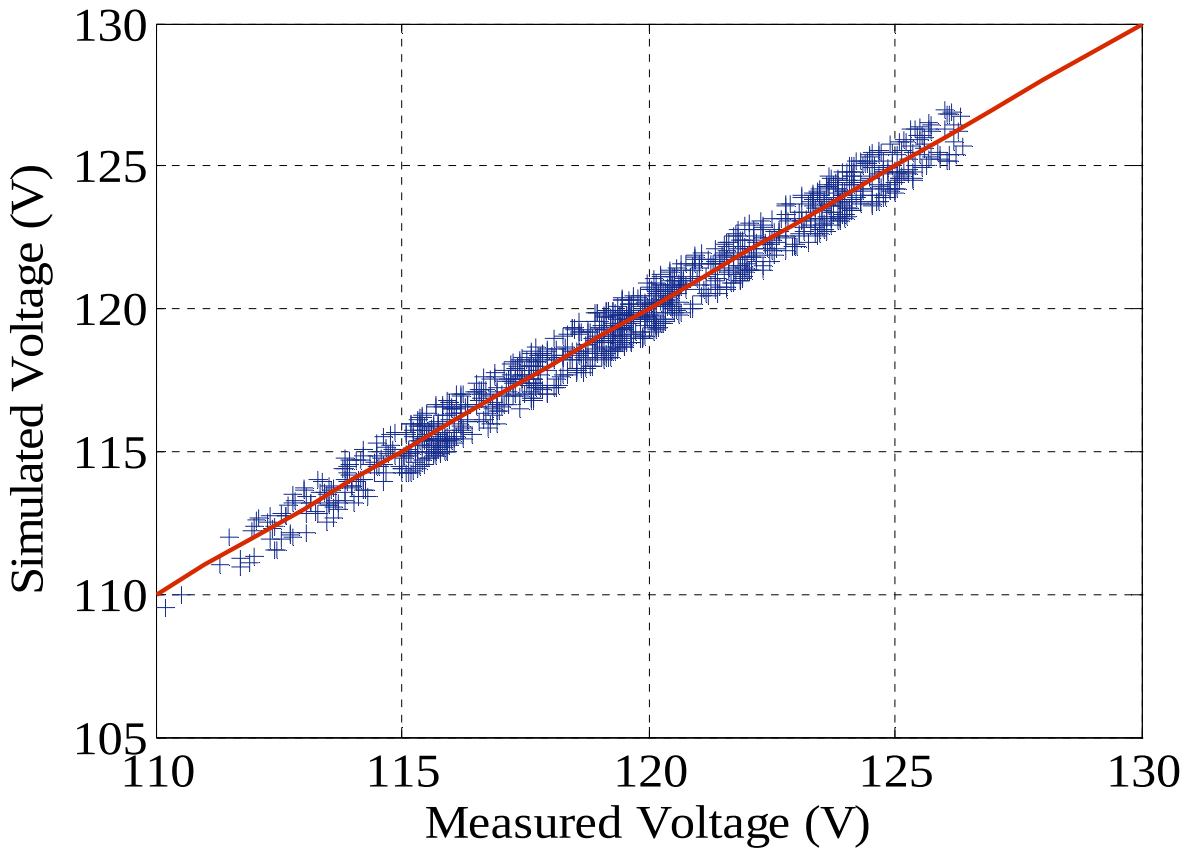


Figure 2.19. Identification of the maximum power point voltage in normal operation conditions.

2.5.5. Faulty sub array versus healthy sub array

Usually, the faults occurring in a photovoltaic system are divided into two classes temporary faults and permanent faults.

2.5.5.1. Temporary faults

Temporary faults like shading may occur due to cloudy weather, snow, dusty and sandy PV array, in addition to building, trees, leaves and bird excrement. Figure 2.20 illustrates the I-V characteristics of PV string with shading effect. While various cases of shading in PV generator are exemplified in Figure 2.21.

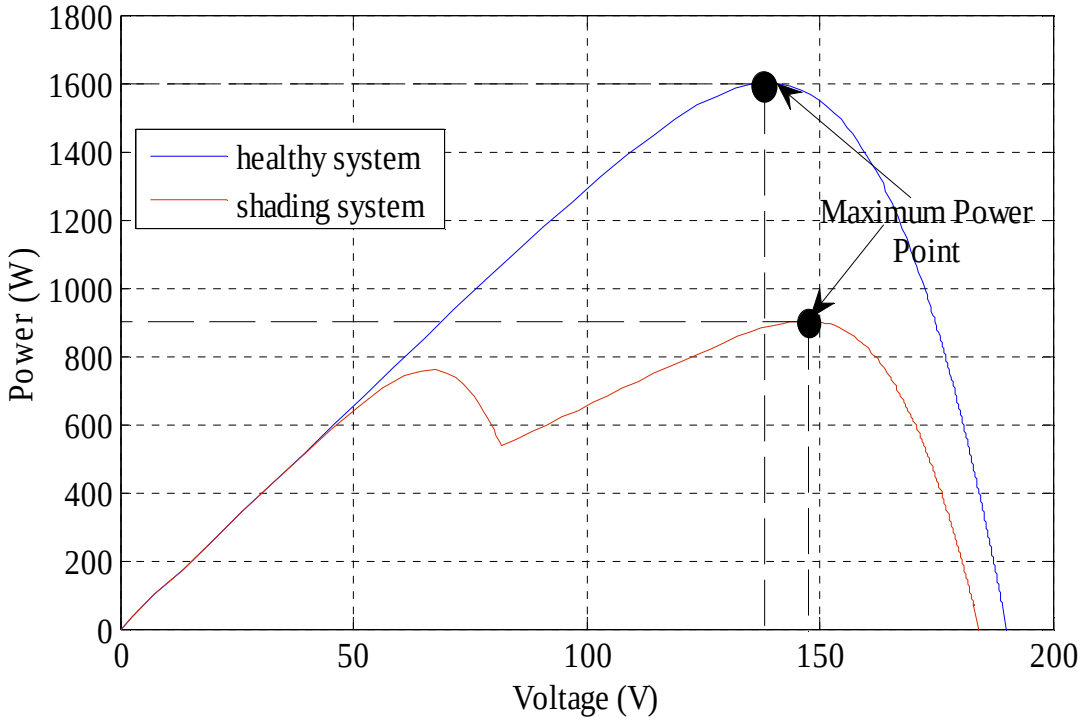
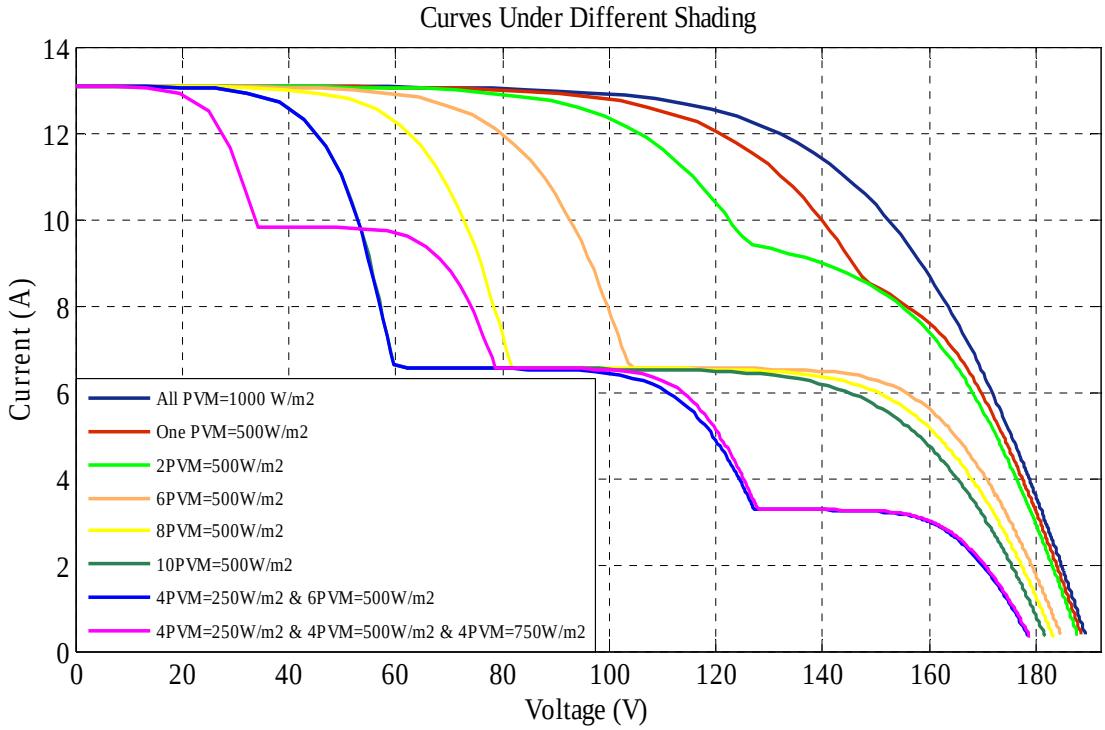
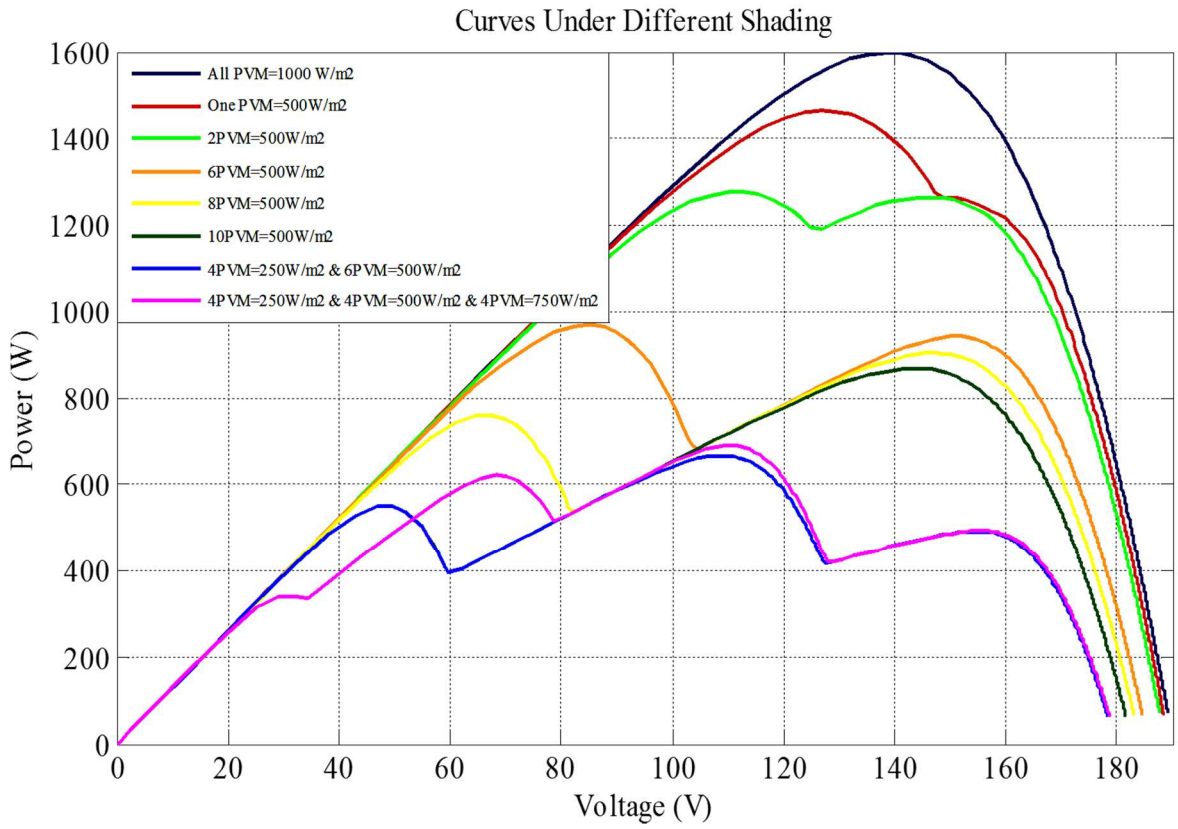


Figure 2.20. Example of temporary PV faults.



(a) Current versus voltage under different shadings.

Figure 2.21. Various examples of shading PV faults (a).



(b) Power versus voltage under different shadings.

Figure 2.21. Various examples of shading PV faults (b).

In our study, those temporary faults are not taken into consideration where the fault is automatically reset after a given period of time and the system returns back to normal operation conditions.

2.5.5.2. Permanent faults

Permanent faults are the most crucial to be analysed notably short circuit and open circuit since they are enduring and ultimately requiring human intervention. The open circuit faults are caused by a sudden disconnection between cells or between modules/strings of the PV array. The I-V characteristics corresponding to these faults can cause important power losses as illustrated by Figure 2.22. These changes at the I-V characteristics shape could therefore be used to identify the corresponding faults.

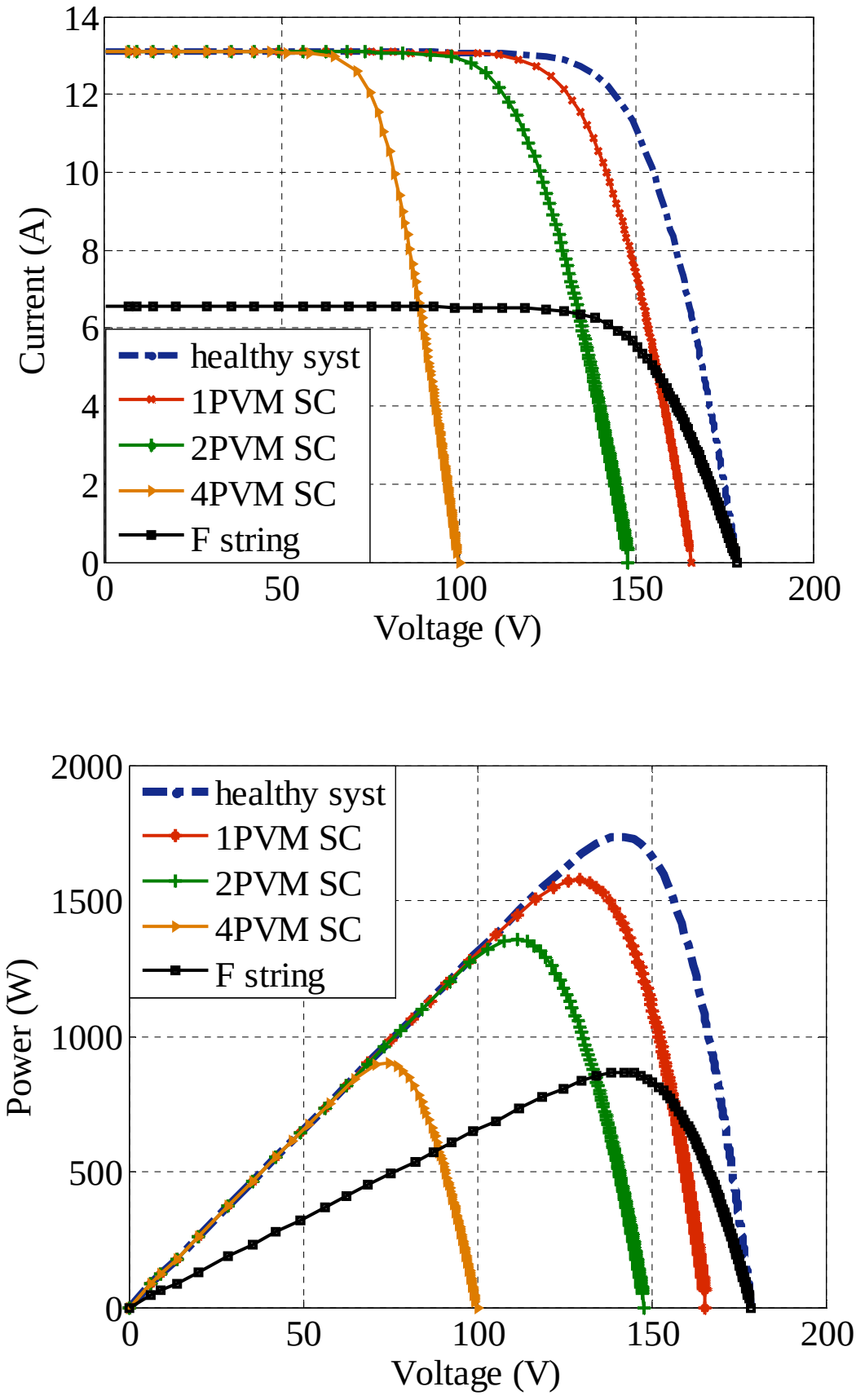


Figure 2.22. Examples of permanents PV faults.

2.5.6. Mathematical analysis of faulty electrical parameters

This developed model has been used to detect the faulty behaviour of the PV system based on the deviation between the desired output and the real output under the same working conditions. This difference between the model output and the system output is the residual information.

2.5.6.1. Faulty current deviation

In Figure 2.23, the studied data show that the string fault leads to a remarkable diminution of the current. For the 1000 analysed samples, the faulty current has been decreased by 50 percent. It should be noted that this remarkable diminution of the current guarantees the corresponding fault detectability and minimize any risk of confusion.

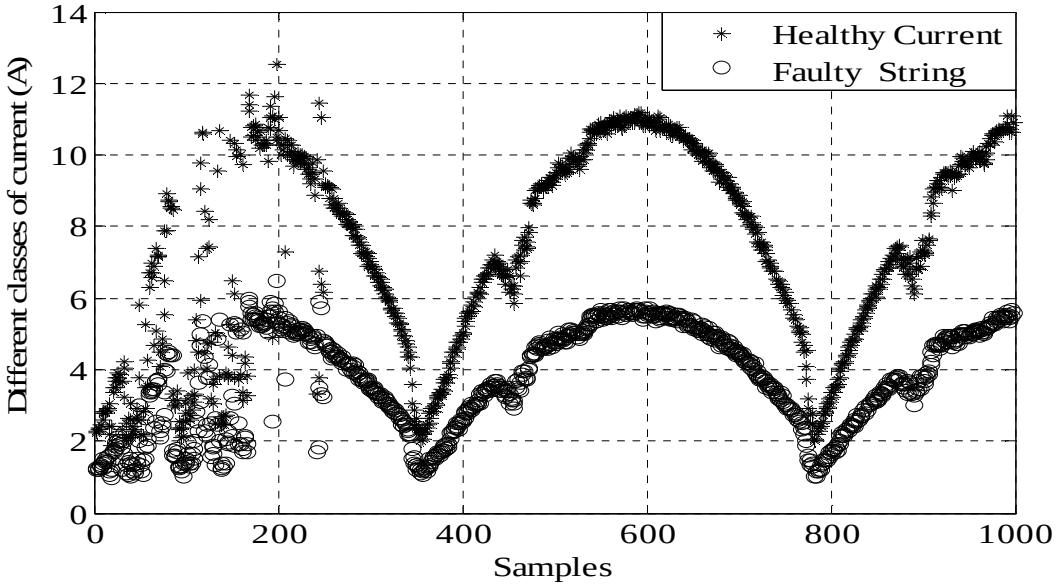


Figure 2.23. The impact of the string fault on the current.

Concerning the current, the faulty system shows a 50 % fall compared to the healthy system as illustrated by Figure 2.24.

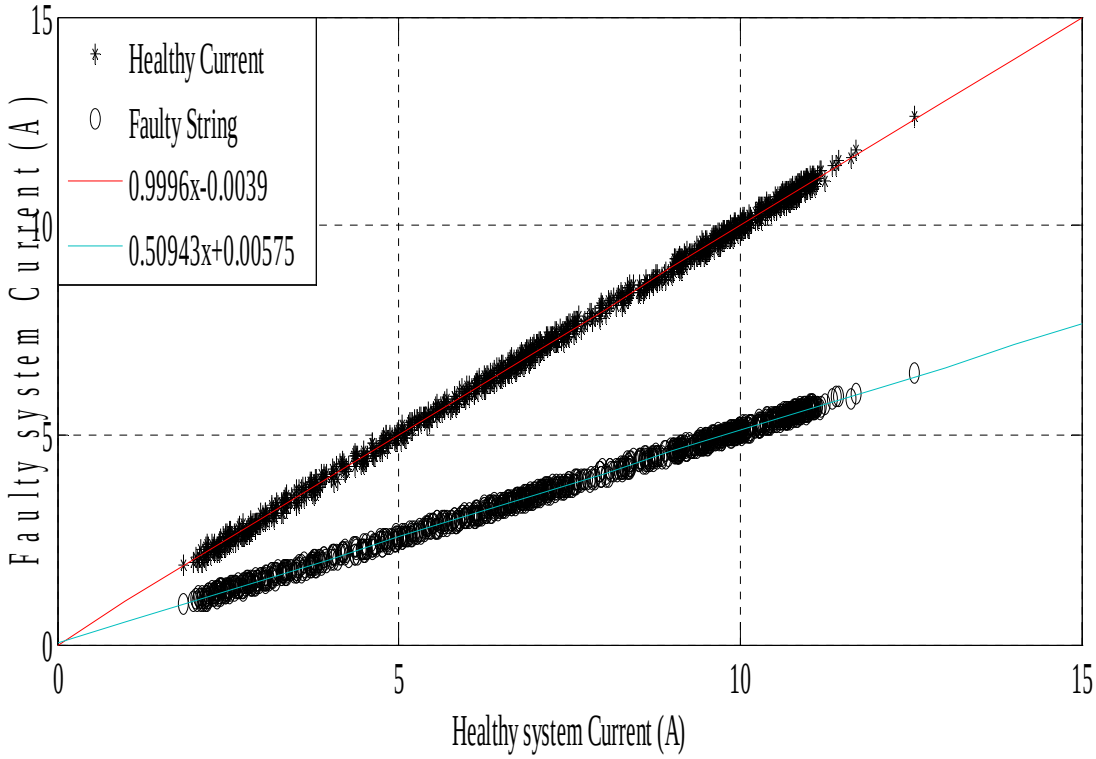


Figure 2.24. Faulty simulated current via healthy measured current of the system.

2.5.6.2. Faulty voltage deviation

As shown in Figure 2.25, which demonstrates that confusion can arise in the overlap area between the two voltages in cases of healthy voltage and one short-circuited PV module. This variation is due to temperature variation, when the temperature increases the voltage decreases and vice versa. In this case, the MPP voltage of the healthy PVG varies from 112V to 127 V, while the faulty PVG voltage (one PV module short circuited) varies between 99V and 116 V.

Figure 2.26 shows that this overlap does not exist when two PV modules are short-circuited in the same string. In this case, the MPP voltage of the faulty PVG varies between in the interval 99V and 116 V. This diminution is more remarkable when four PV modules are short-circuited in the same string. As illustrated by Figure 2.27, the MPP voltage of the faulty system varies between 65V and 74 V. This close correlation between the variation's interval of the faulty PVG voltage and the number of the short-

circuited modules can to be used to detect and identify the fault as illustrated by Figure 2.28.

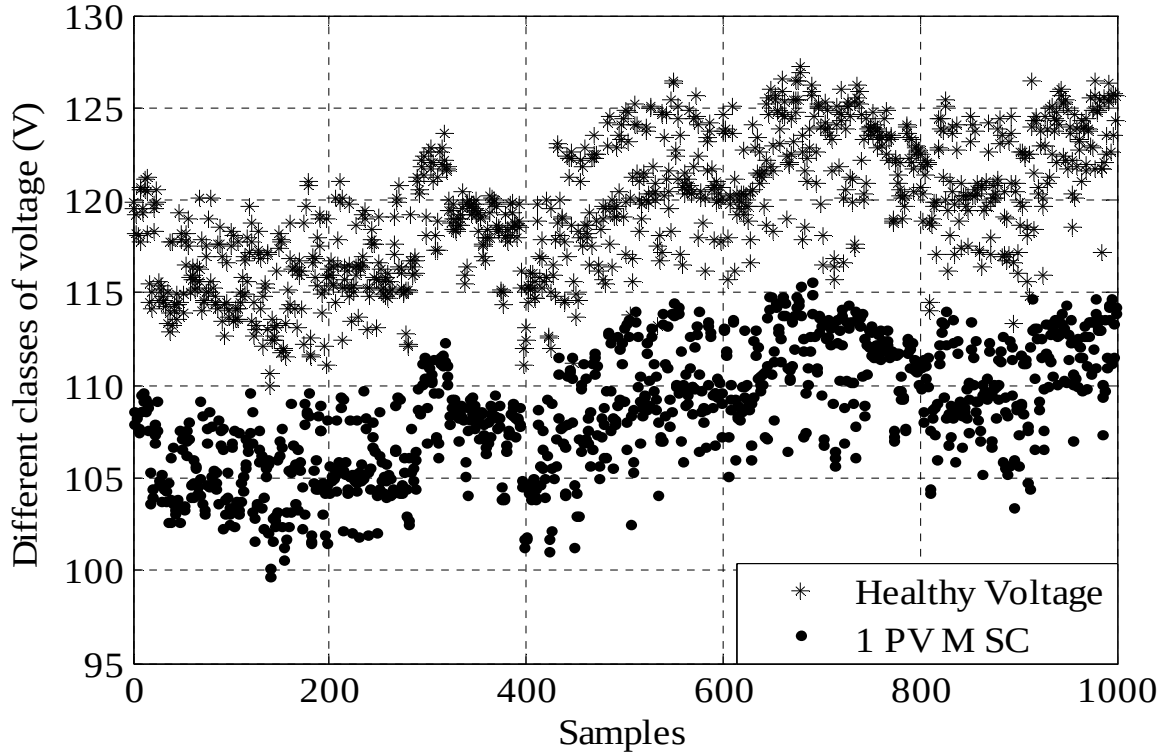


Figure 2.25. The voltage of a PVG with one short-circuited module.

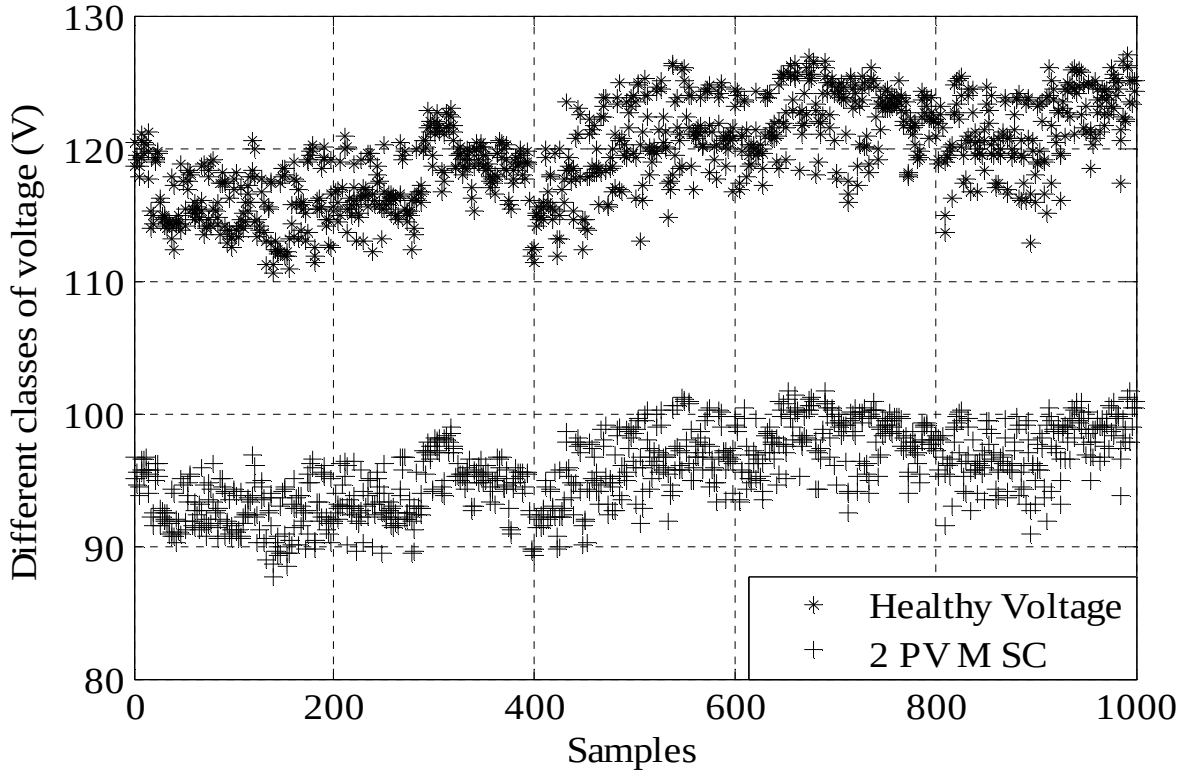


Figure 2.26. The voltage of a PVG with two short-circuited modules.

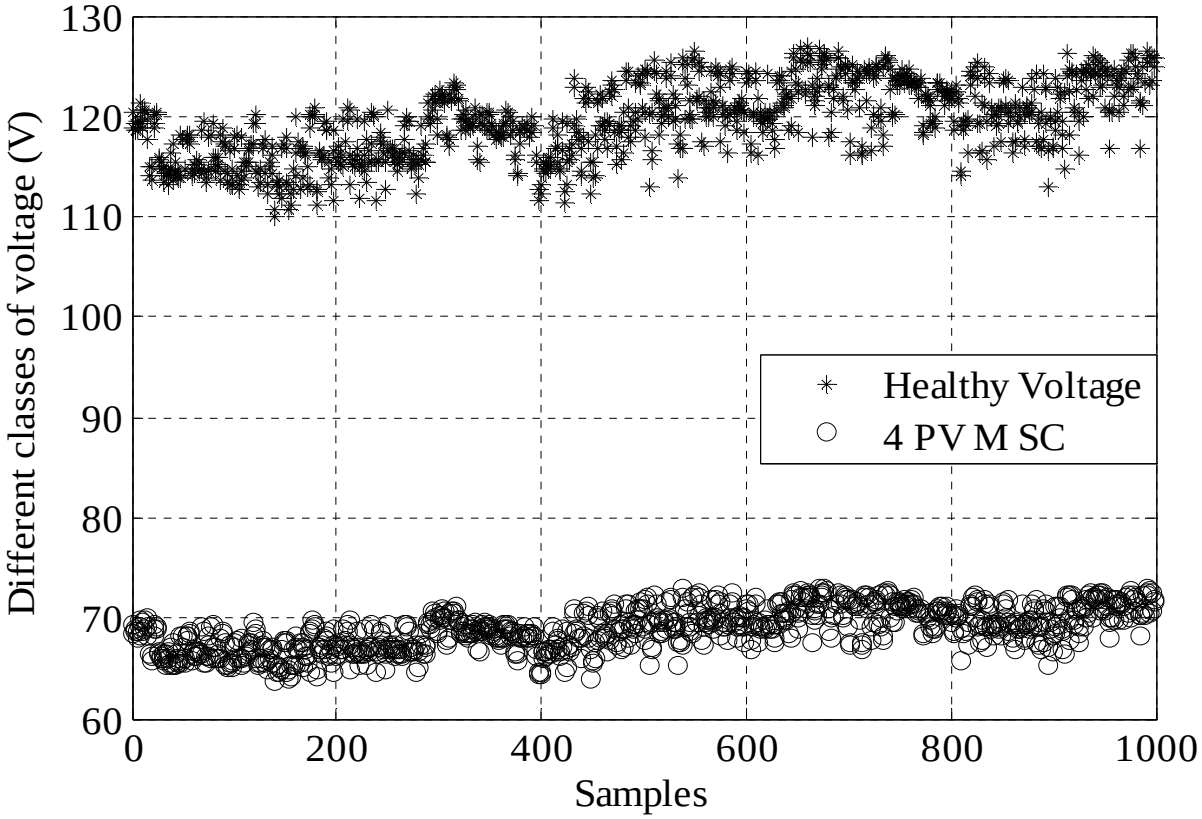


Figure 2.27. The voltage of a PVG with four short-circuited modules.

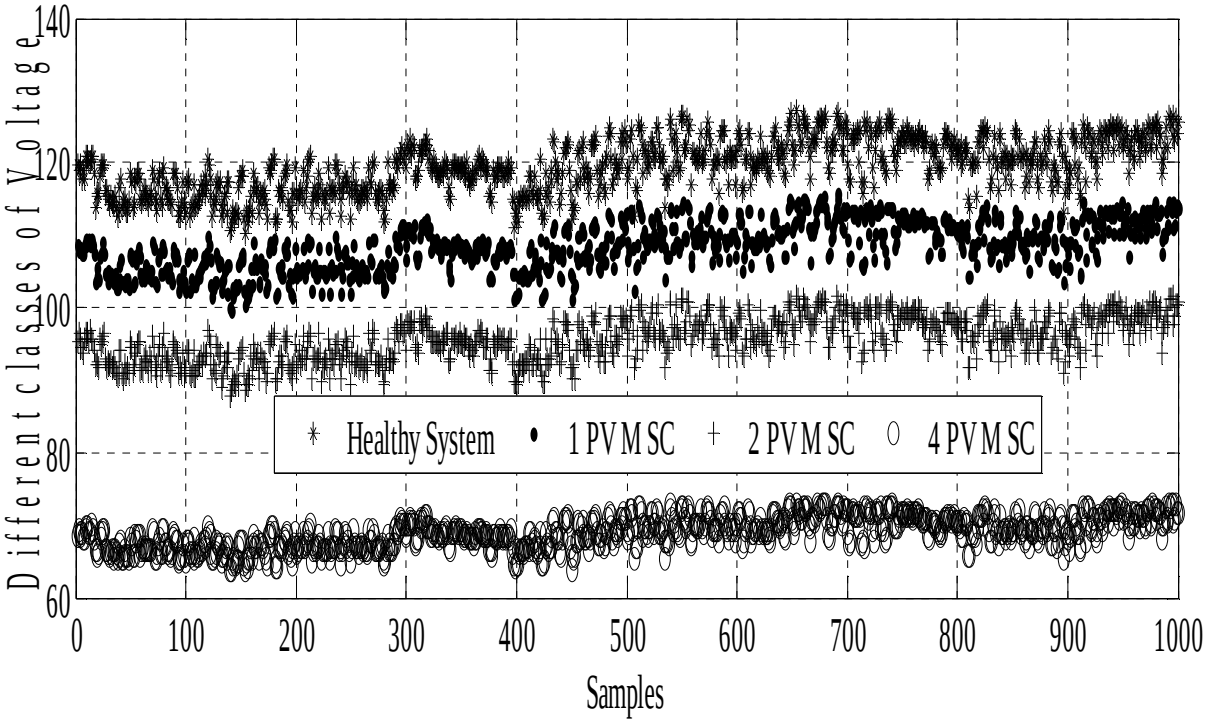


Figure 2.28. The voltage of global different short-circuited modules in PVG.

The comparison between these three faults classes and the healthy case has been used to establish the correlation illustrated by Figure 2.29. The PVG have a 10% lower voltage when only PV module is short-circuited, 25% and 48% lower voltages when two and four short-PV modules are short-circuited respectively.

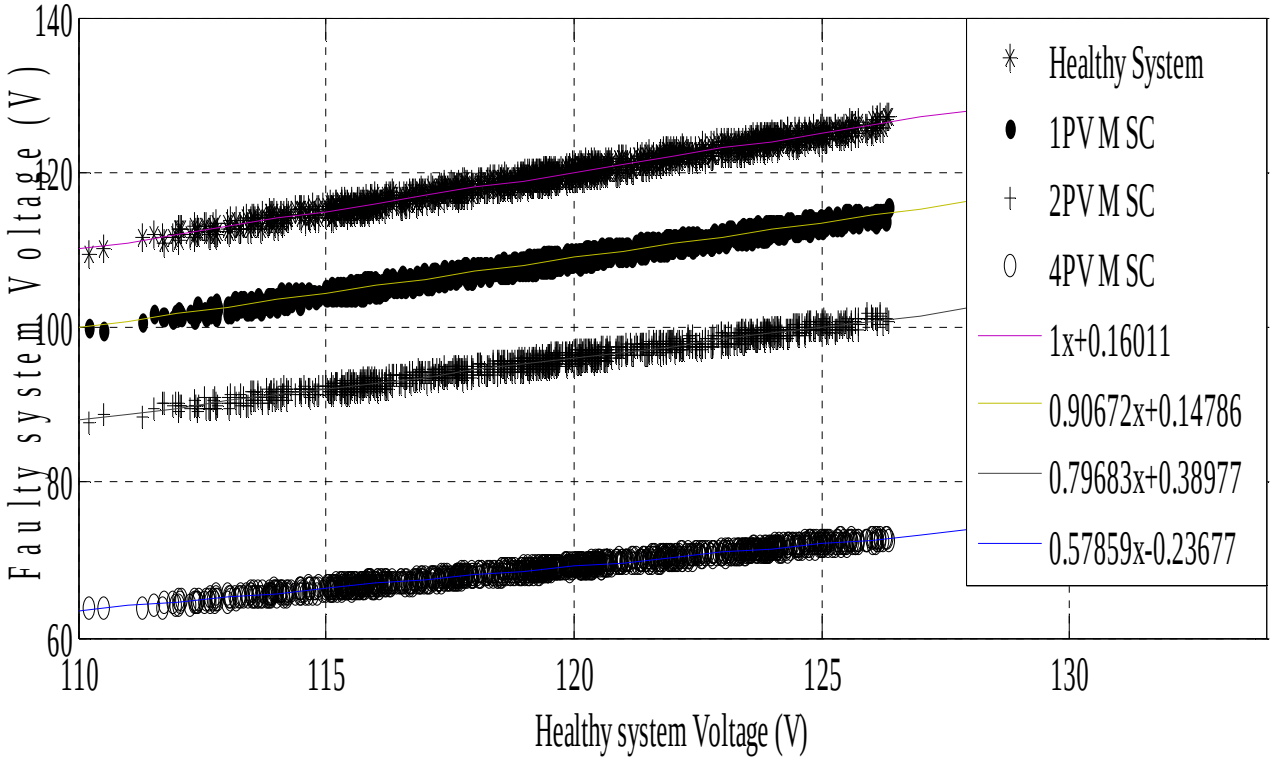


Figure 2.29. Faulty simulated voltage via healthy measured voltage of the system.

For the fault identification, the IFD algorithm combines the voltage residual information and the current residual information. This mixed information enables to build the following correspondence table.

Table 2.3. Faults correspondence table.

	Impp (%)	***Vmpp (%)***	***Pmpp (%)***
Healthy system	100	100	100
1 PV M SC	↘ 1	↘ 10	↘ 11
2 PV M SC	↘ 1	↘ 25	↘ 26.5
4 PV M SC	↘ 1	↘ 48	↘ 49
Faulty string	↘ 50	100	↘ 50

2.6. Conclusion

This chapter presents the experimental PV plant with its different components followed by PV array modelling study with PV cell modeling presentation as well as I-V characteristic of PV array under variation of working conditions. Then after, a creation of simulated PV array using Simpscape / MATLAB Simulink identical to real experimental PV array has been presented where the comparative study should be discussed in order to pay attention on the efficiency and precision by residual computations. The third stage is considered as the interest's center of our study in order to elaborate a database with healthy operation and different faults encountered in PV generator such as open circuit fault and various cases of short circuit citing one PV module short circuited, two PV modules short circuited and four PV modules short circuited. The final stage of this chapter has been consecrated to the mathematical analyze of different faults comparing to healthy operation with both electrical parameters current and voltage of maximal power point.

CHAPTER 3 FAULT DIAGNOSIS OF PHOTOVOLTAIC GENERATOR USING BACK PROPAGATION NEURAL NETWORK.

3.1. Introduction

This chapter presents a new intelligent algorithm for PV system's diagnosis and fault detection (IFD). This approach requires algorithm that can detect and identify three recurrent cases between, healthy, short circuit fault as well as disconnected branch in PV generator using two Back Propagation Neural Networks (BPNN). The first ANN is deduced for I_{mpp} classification and the second ANN for the V_{mpp} classification. Both detection and isolation are simple and fast. The developed model requires small training period and is based on only four inputs: the maximal power current and voltage from the output I-V characteristic, the solar irradiation and the cell temperature.

The follow steps are consecrated to give explanation and more detailed about this approach as well as its efficiency on fault detection in PV systems.

3.2. Methodology

As described by Figure 3.1, the implementation of the present IFD is as a three-steps procedure: 1) The detection phase demonstrated in detail in chapter 2. 2) The isolation phase: during this stage, two ANNs' classifiers are necessitated to classify the information about the fault which caused the detected residual information. These two ANNs have been already trained to separately classify the current and the voltage at the MPP. 3) The identification phase: during this final stage, the output of the ANNs is analysed by a combinational algorithm called the logical block whose role is to identify and recognize the corresponding fault.

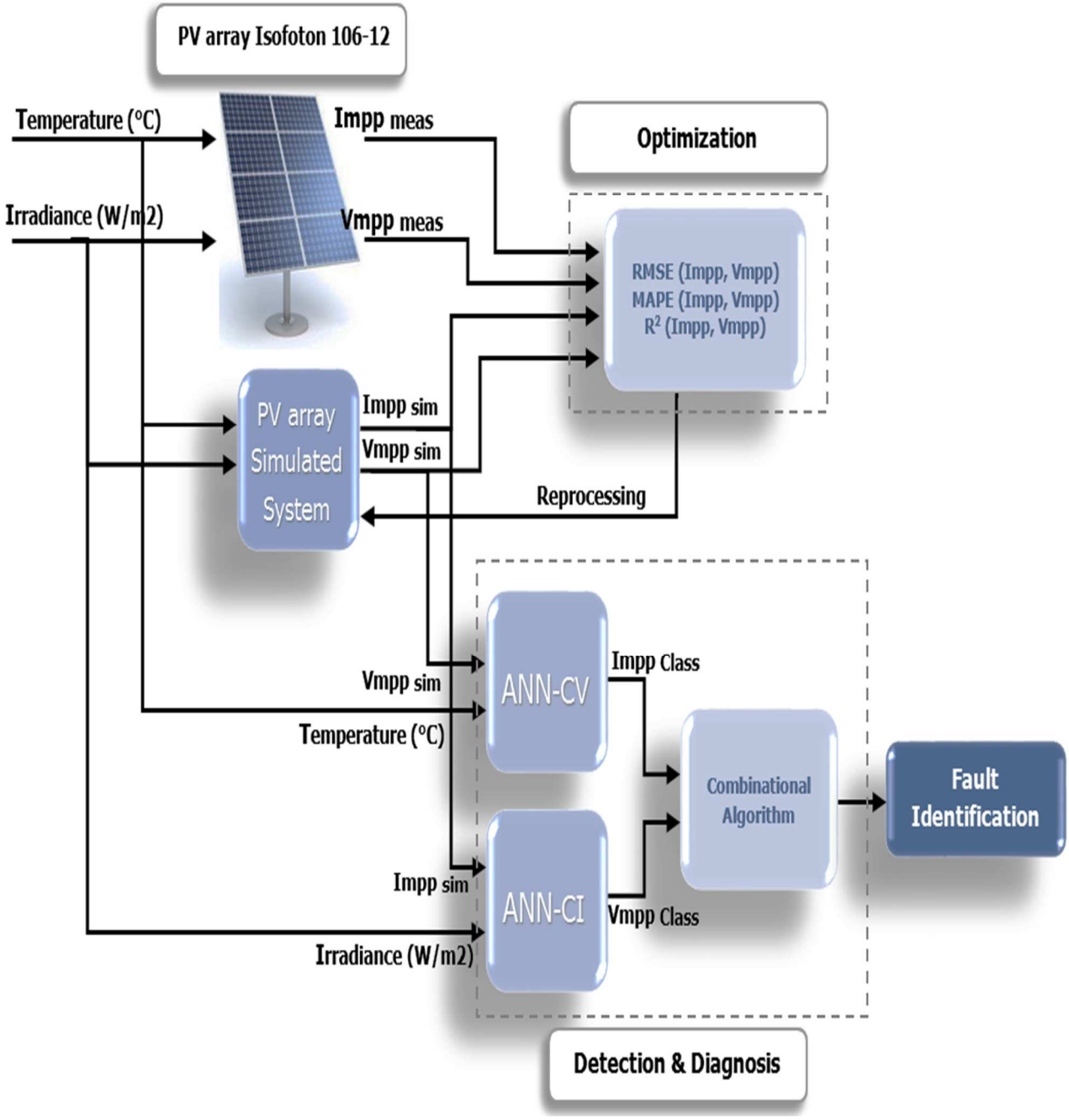


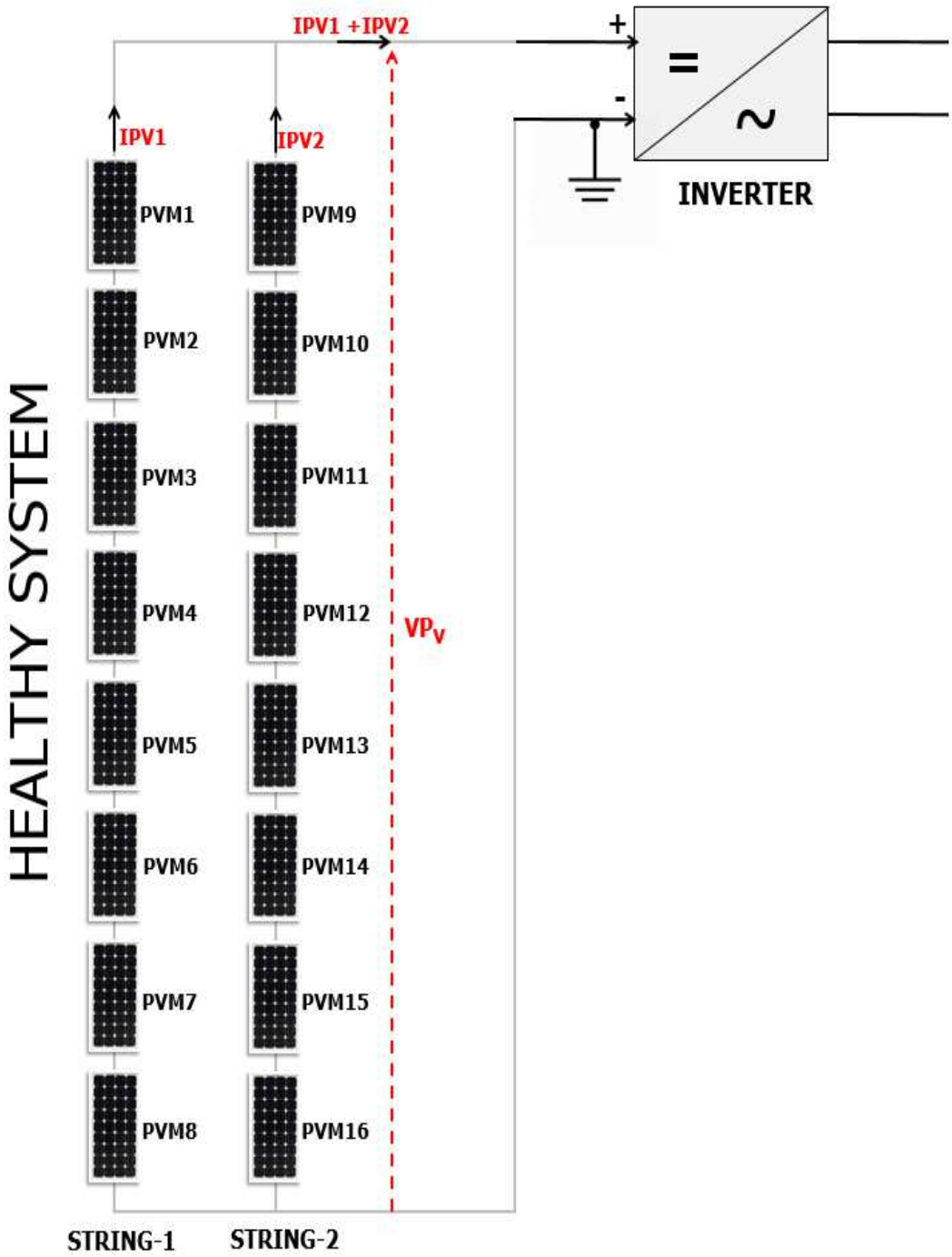
Figure 3.1. Schematic description of the IFD methodology.

3.3. IFD algorithm's description

3.3.1. The various faults treated in this study

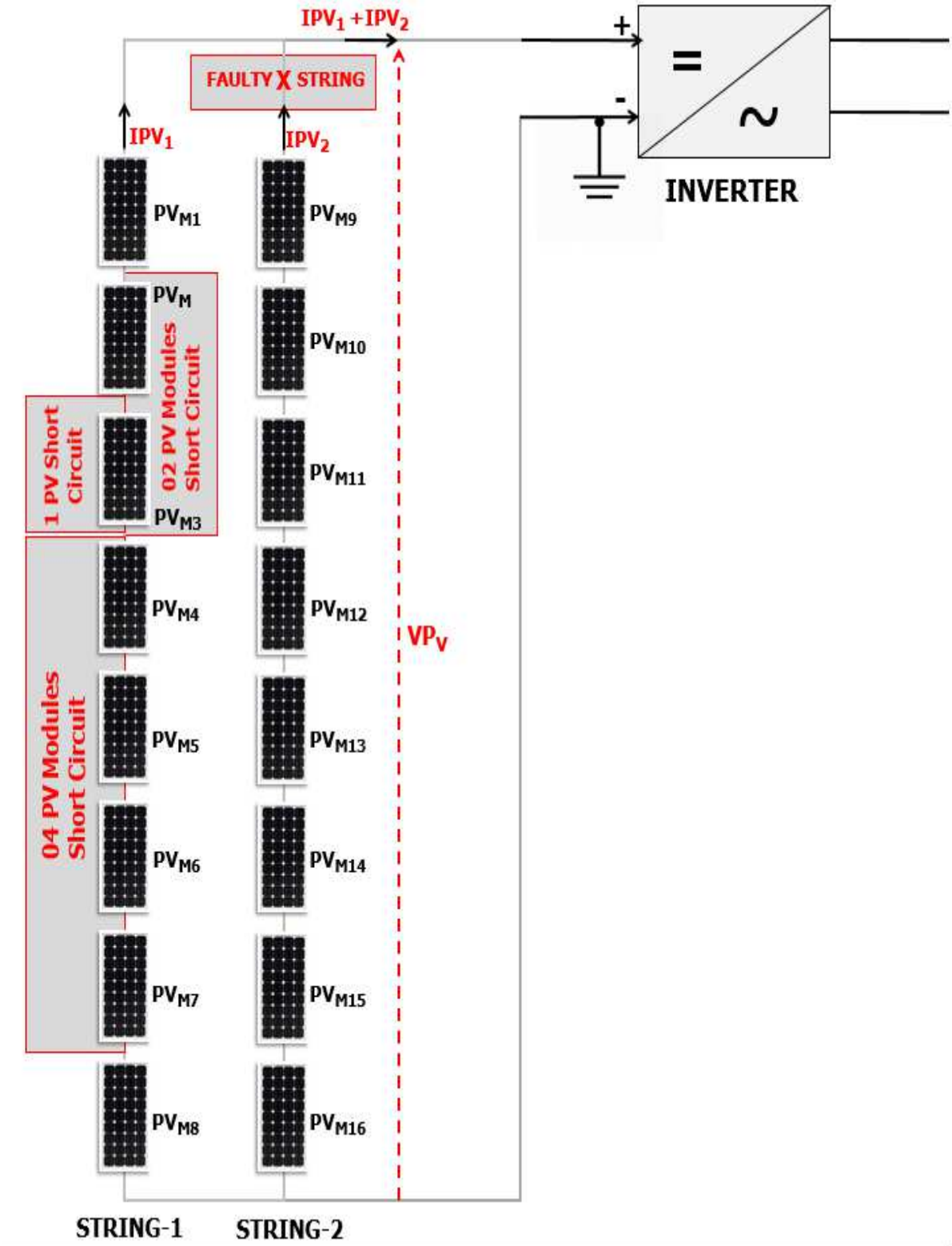
The IFD algorithm is designed to detect, identify and isolate four faulty modes: 1) one PV module short circuit, 2) two PV modules short circuit, 3) four PV modules

short circuit and 4) faulty string. These faults are illustrated by Figure 3.2 and summarized as well as symbolized by Table 3.1.



(1)

Figure 3.2. Overall scheme of the PV system (a) normal operation conditions (b) faulty string and short circuits faults.



(2)

Figure 3.2. Overall scheme of the PV system (a) normal operation conditions (b) faulty string and short circuits faults (the following).

Table 3.1. The different state of the system with faults and their symbols.

Fault	Symbol
Normal operation conditions	C1
Fault detection refers to one panel short circuit in PV string	C2
Fault detection refers to two panels short circuit in PV string	C3
Fault detection refers to four panels short circuit in PV string	C4
Fault detection refers to disconnection string	C5

3.3.2. The ANNs architecture

3.3.2.1. The classifier ANN of Impp

The classification of I_{mpp} requires ANN containing two neurons in its input layer, the first neurons represents the solar irradiation data and the second neuron represents the current data at maximal power point (I_{mpp}). This ANN own one neuron in its output layer devotes for the classification and localization of current. Between both layers, it exists two hidden layers including eight neurons for each one. Figure 3.3 illustrates the architecture of ANN current.

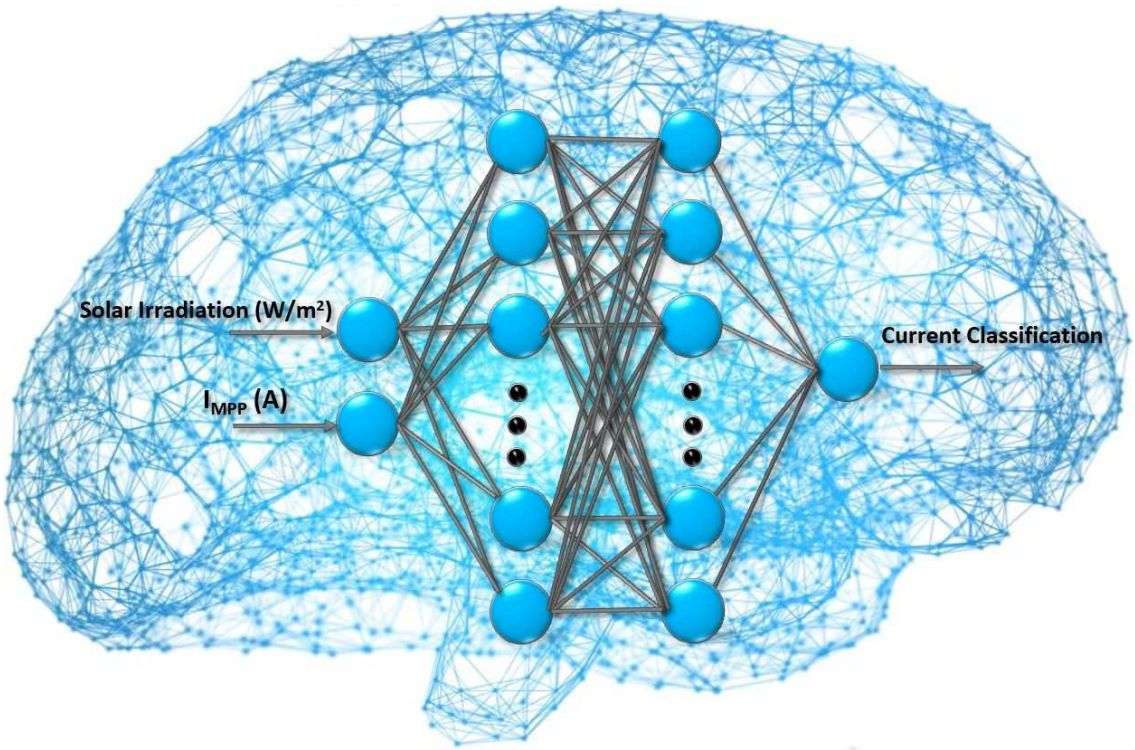


Figure 3.3. Architecture of ANN current.

3.3.2.2. The classifier ANN of Vmpp

This ANN illustrated in Figure.3.4 is deduced for the classification of the voltage in PV generator. The present ANN requires two neurons in its input layer: the first one is reserved for cell temperature data and the second neurons is reserved for voltage data at maximal power point (V_{mpp}) followed by the second layer that represents the hidden layer containing forty five neurons, and finally one neuron in output layer designated for the voltage classification.

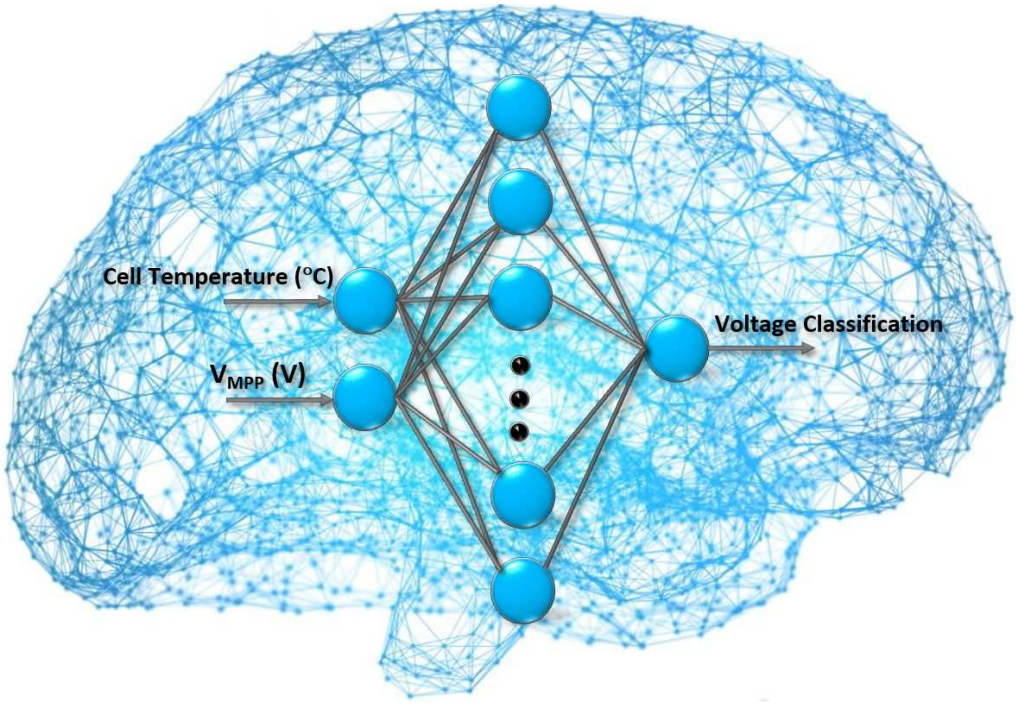


Figure 3.4. Architecture of ANN current.

3.4. Diagnosis procedure in PV generator

The diagnosis is based on a two-stage procedure. In the first stage, the fault is detected by comparing the PV system's and its mathematical model outputs. In case of fault detection, the identification stage is lunched. In both stage ANNs are used. They are exploited to simulate the PV system in the first stage and for the fault classification during the second stage. The classification is based on two ANNs, the first one (ANN-CI) is used for the I_{mpp} classification and the second's (ANN-CV) for the V_{mpp} classification. For the identification stage, a combinational algorithm is used to analyse the findings of the classification stage, which are based on the I_{mpp} and V_{mpp} values

(codes). The different classes and their combinations are summarized in Tables 3.2 and 3.3 respectively.

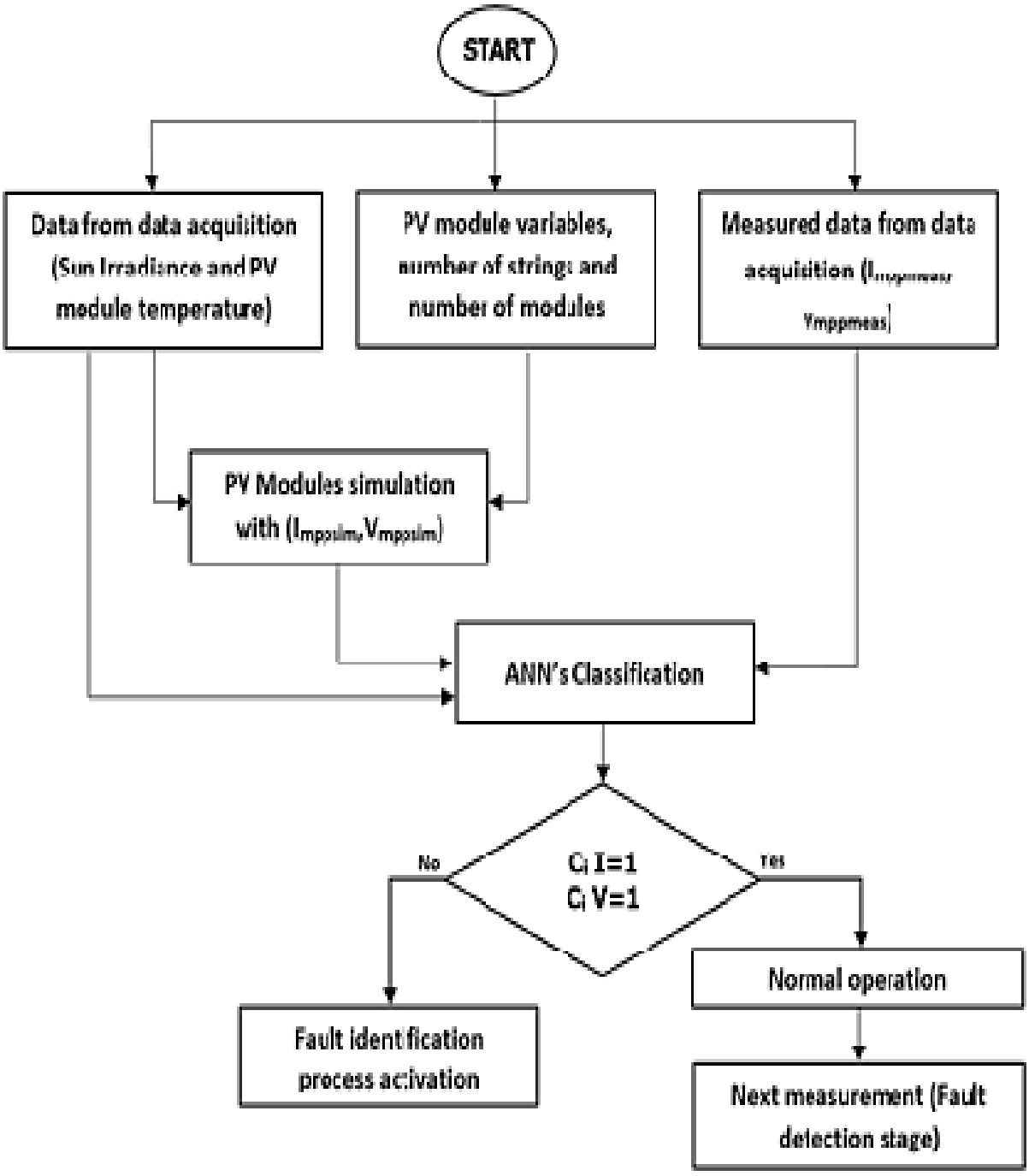
Table 3.2. Classifications stage, classes and symbols.

Symbols	Description	Classes
Impp n	Maximum power point current of normal operation conditions	Class1I
Impp str	Maximum power point current of faulty string	Class2I
Vmpp n	Maximum power point voltage of normal operation conditions	Class1V
Vmpp 1sc	Maximum power point voltage of one module short circuit	Class2V
Vmpp 2sc	Maximum power point voltage of two modules short circuit	Class3V
Vmpp 4sc	Maximum power point voltage of four modules short circuit	Class4V

Table 3.3. Different combination of classes obtained.

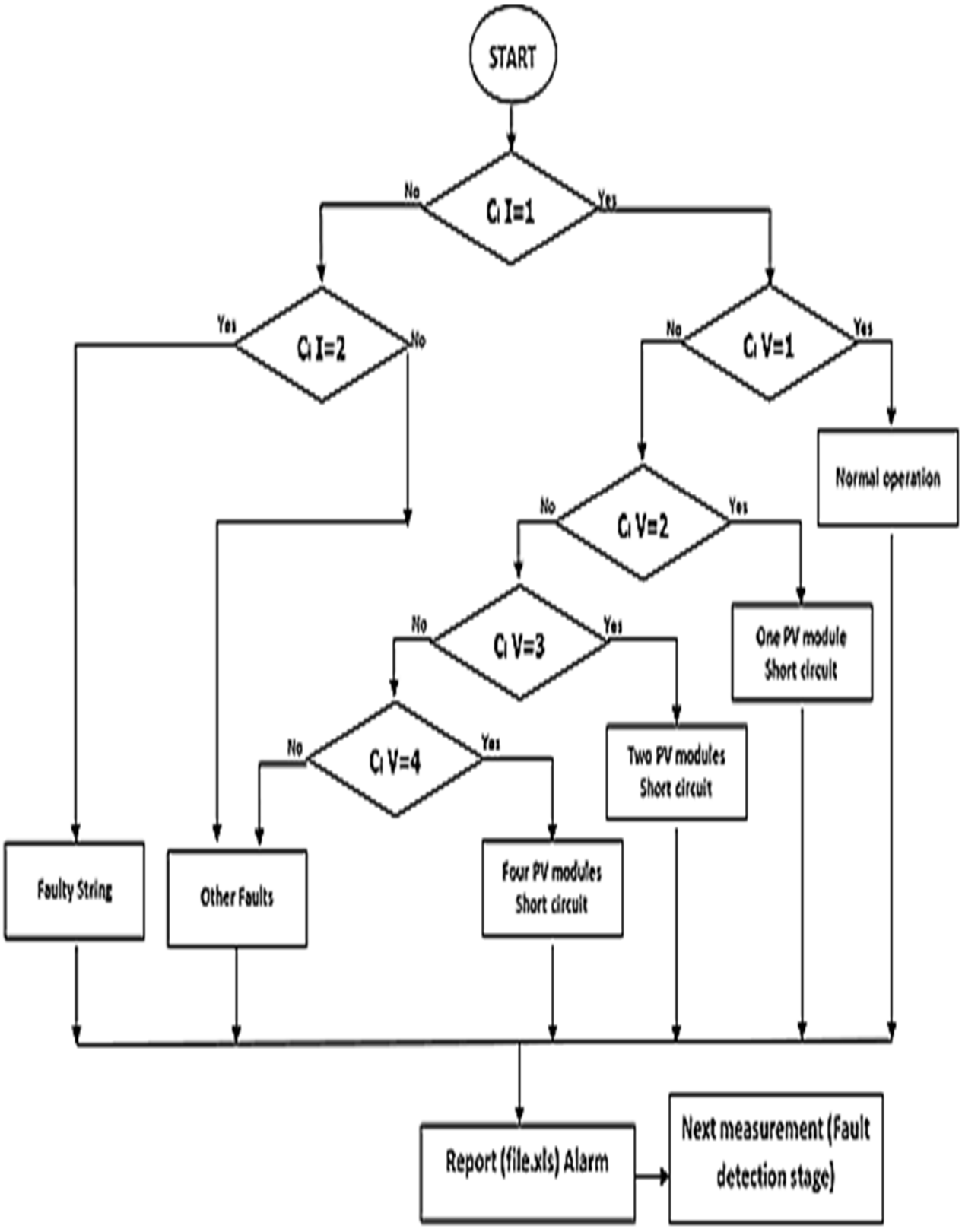
Impp	Impp Class	Vmpp	Vmpp Class	Global Description	Global Classification
Impp n	Class1I	Vmpp n	Class1V	Normal operation conditions	Class1
Impp n	Class1I	Vmpp1sc	Class2V	One faulty PV module in string	Class2
Impp n	Class1I	Vmpp2sc	Class3V	Two faulty PV modules in string	Class3
Impp n	Class1I	Vmpp4sc	Class4V	Four faulty PV modules in string	Class4
Impp str	Class2I	Vmpp n	Class1V	Faulty string	Class5

Figure 3.5 demonstrates clearly two flowcharts indicating the stages in this study. The first flowchart shows the first steps procedure starting by the collect of wheather data from renewable energy station as well as the electrical measured and simulated data to ANN detection, where if both codes ANNs display 1 mean that the model is in its normal operation else the activation of fault identification process is necessary. On the other hand, the second flowchart explains precisely the code process for both ANN to identify the kind of fault encounter in the PV generator.



a) Fault detection stage.

Figure 3.5. Flowchart of the IFD algorithm.



b) Fault identification stage

Figure 3.5. Flowchart of the IFD algorithm (the following).

The end of this section consists to simulate the diagnosis of the model using two ANN's trained and tested, the output of both ANNs represents the input of combinational algorithm in order to obtain the final classification. The obtained database presents 10000 samples of each attribute, divided in two parts: 5000 samples for the ANN classification current and 5000 samples for the ANN classification voltage.

The first ANN classifies the current of the maximum power point. For the training phase, 2000 samples have been used. Each sample contains the current value at the MPP (Impp(A)) and the irradiance level (W/m^2) as input data. 50% of the samples represent healthy operation conditions and remaining 50% represent the disconnected string with the combination summarized in Table 3.4.

Table 3.4. Faults correspondence table.

Faults	Number of samples for each attribute
Healthy current	50 samples for each attribute.
Four PV module short circuited	40 samples for each attribute.
One PV module short circuited	100 samples for each attribute.
Two PV module short circuited	52 samples for each attribute.
Healthy voltage	30 samples for each attribute.
Two PV module short circuited	70 samples for each attribute.
One PV module short circuited	20 samples for each attribute.
Four PV module short circuited	38 samples for each attribute.
Healthy current	11 samples for each attribute.
Disconnected string	58 samples for each attribute.
Healthy current	31 samples for each attribute.

Table 3.5. Faults correspondence table.

Faults	Number of samples for each attribute
Healthy voltage	50 samples for each attribute.
Four PV module short circuited	40 samples for each attribute.
One PV module short circuited	100 samples for each attribute.
Two PV module short circuited	52 samples for each attribute.
Healthy voltage	30 samples for each attribute.
Two PV module short circuited	70 samples for each attribute.
One PV module short circuited	20 samples for each attribute.
Four PV module short circuited	38 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Disconnected string	58 samples for each attribute.
Healthy voltage	31 samples for each attribute.

The second ANN requires in its training, 2000 samples have been used for each attribute (Vmpp (V), Temperature (°C)) it means (2000 x 2= 4000 data) as input data, for training phase; 25% samples represent healthy voltage (500 samples for each attribute) and the remaining 75% samples represent different faults divided into three equal categories which are: 25% for one PV module short circuited, 25% for two PV modules short circuited and 25% for four PV modules short circuited. For the ANN diagnosis 500 samples has been employed of each attribute (Vmpp (V), Temperature (°C)) distributed in eleven cases respectively as summarized in Table 3.5.

3.5. Simulation Results and interpretations

This section is consecrated for the behavior of both ANN, their classification under different faults, the global diagnosis of the PV generator as well as the performance of the system studied.

3.5.1. Behavior of ANNs in PV generator fault diagnosis

In this sub-section, the classification of each electrical parameter has been analyzed separately.

3.5.1.1. Current of maximal Power Point (Impp)

a- Training phase

The training stage of this ANN requires Marquardt Levenberg algorithm. In order to obtain a good quality from point of view of classification, this ANN necessitates 2800 iterations and a performance equal to 0.0001 during 7:52min shown in Figure 3.6.

The curve of ANN's training Error of Current shown in figure 3.7 decreases from 10 and stabilized at 0.007 that indicates the precision of the model. In Figure 3.8 the samples are in their right classes and the coefficient of regression displays 0.9808 value, which proves a top quality of classification.

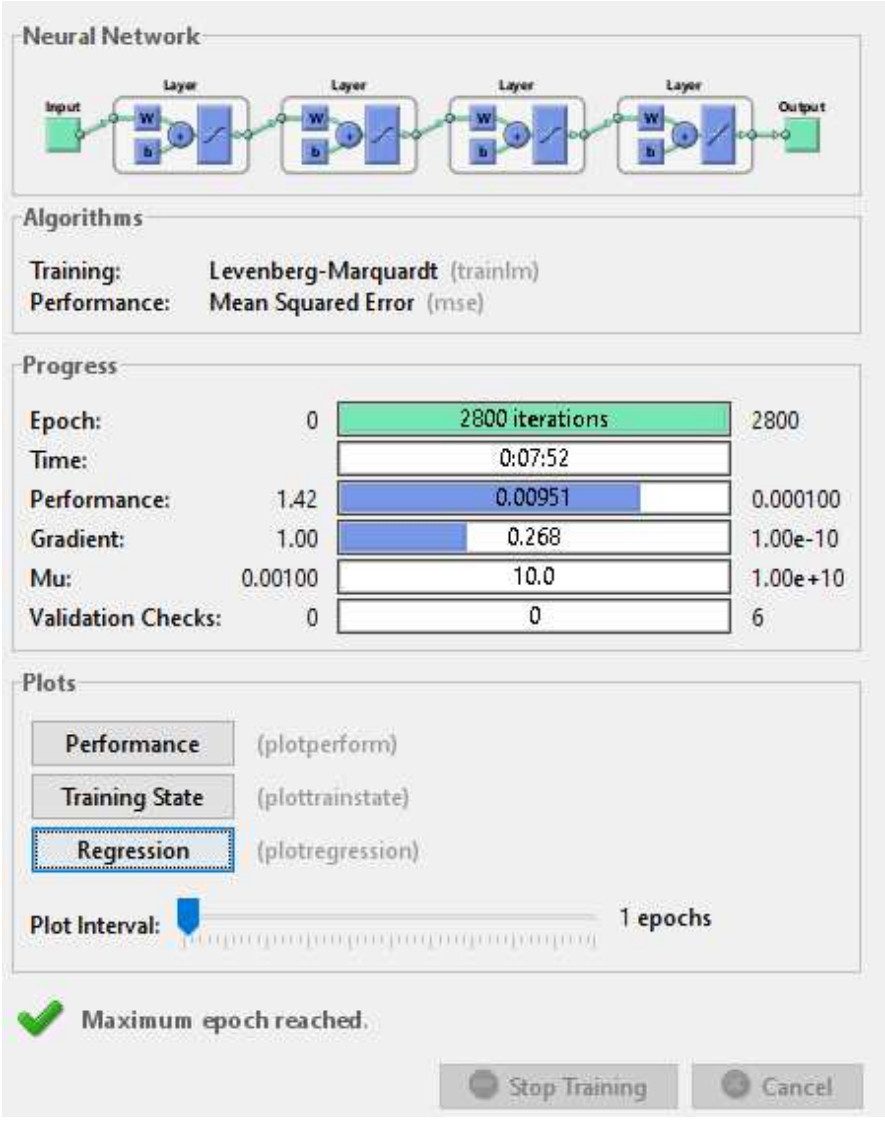


Figure 3.6. ANN's training phase of Current.

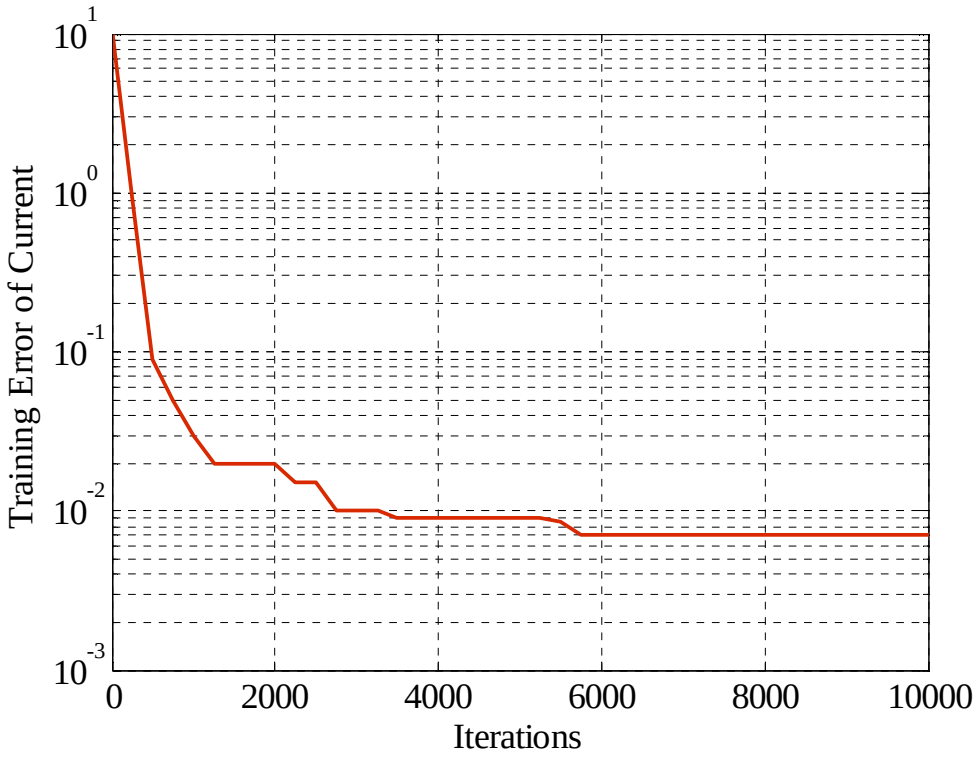


Figure 3.7. ANN's training Error of Current.

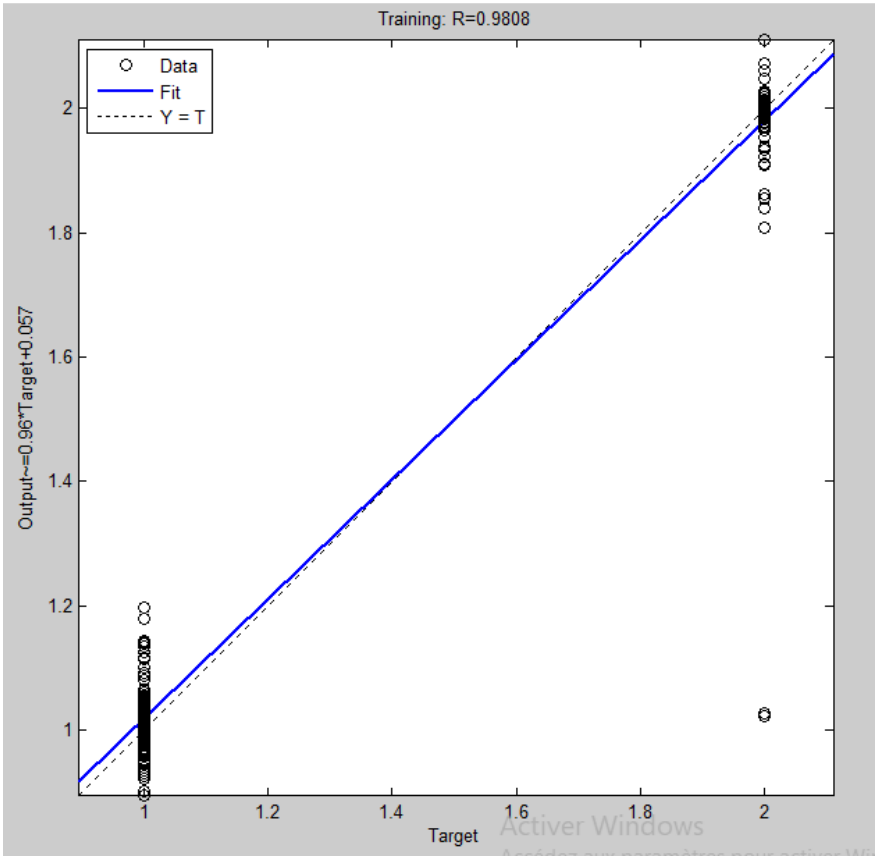


Figure 3.8. Classification accuracy of Voltage.

b- Validation phase

Figures 3.9 illustrates the classification results for the ANN current. This shows the ability of the obtained ANNs to introduce the incoming data in their right classes with high accuracy; 99.6% for the current classification. The accuracy of the current classification is very important in this stage in order to obtain a good result of diagnosis.

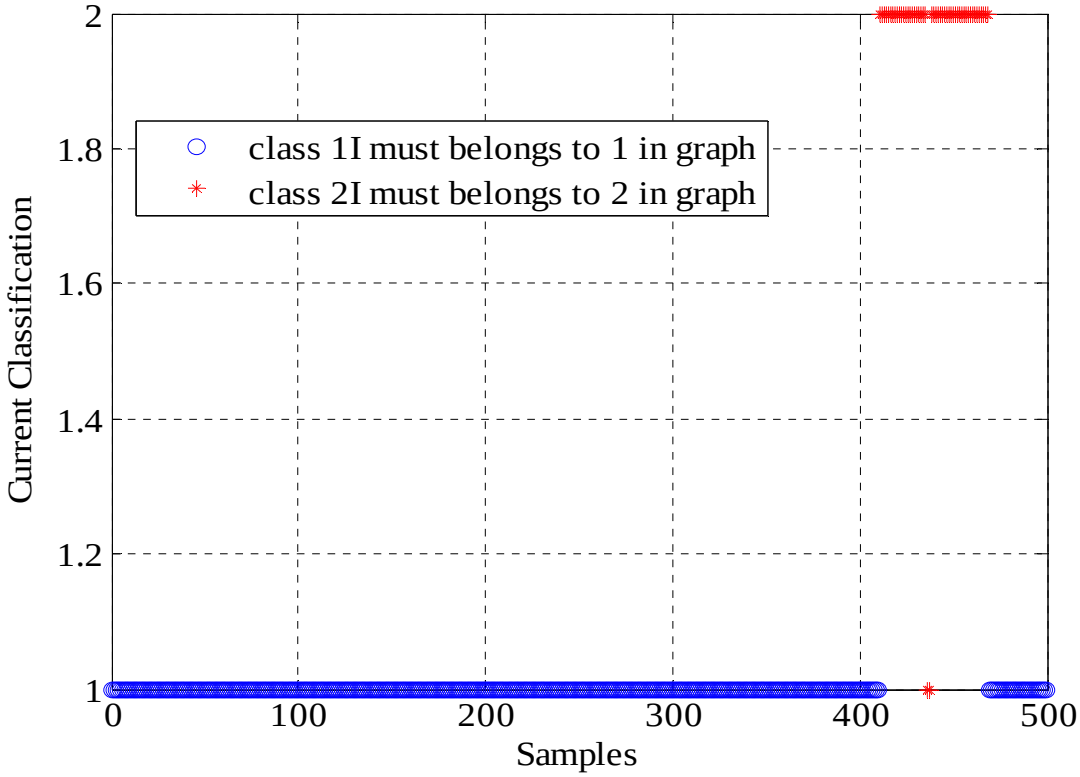


Figure 3.9. Impp classification using ANN.

3.5.1.2. Voltage of maximal Power Point (Vmpp)

a- training phase

As shown in Figure 3.10. the training stage of this ANN requires Marquardt levenberg algorithm. In order to obtain a good classification, the number of iterations is 1750; its performance is equal to 10^{-5} during 6:43 minutes. The figure 3.11 demonstrates the training error of the voltage issue from the ANN, where the decrease of curve is clear going from 10 to 0.007 that proves the precision and efficiency of the classification. The figure 2.12 demonstrates the classification under linear curve displaying the coefficient of regression, which is 0.99934 that proves a high level of pre-

cision from point of view of classification, where the figure demonstrates that the majority of samples are in their right classes, the only confusions are with three between class1 and class 2.

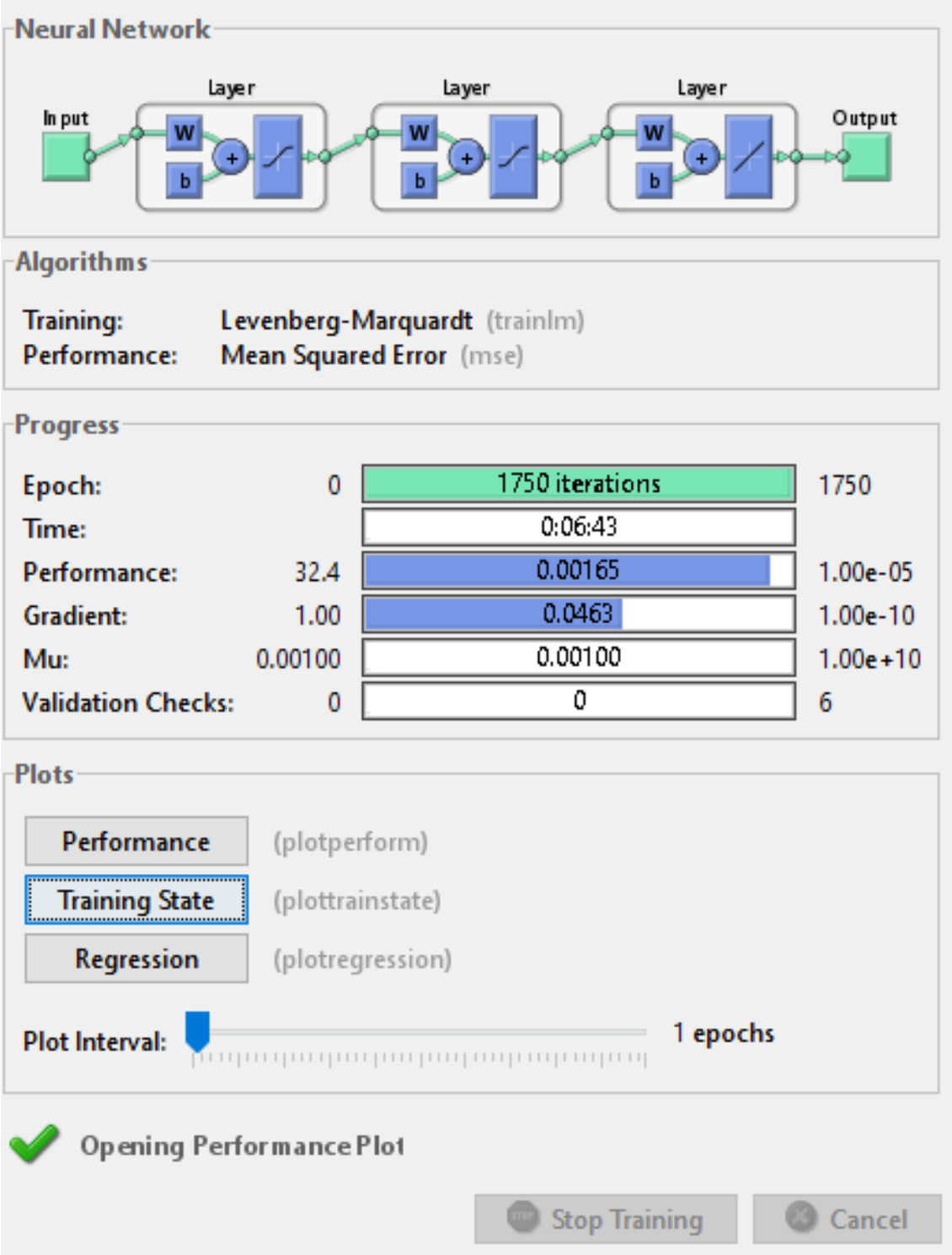


Figure 3.10. ANN's training phase of Voltage.

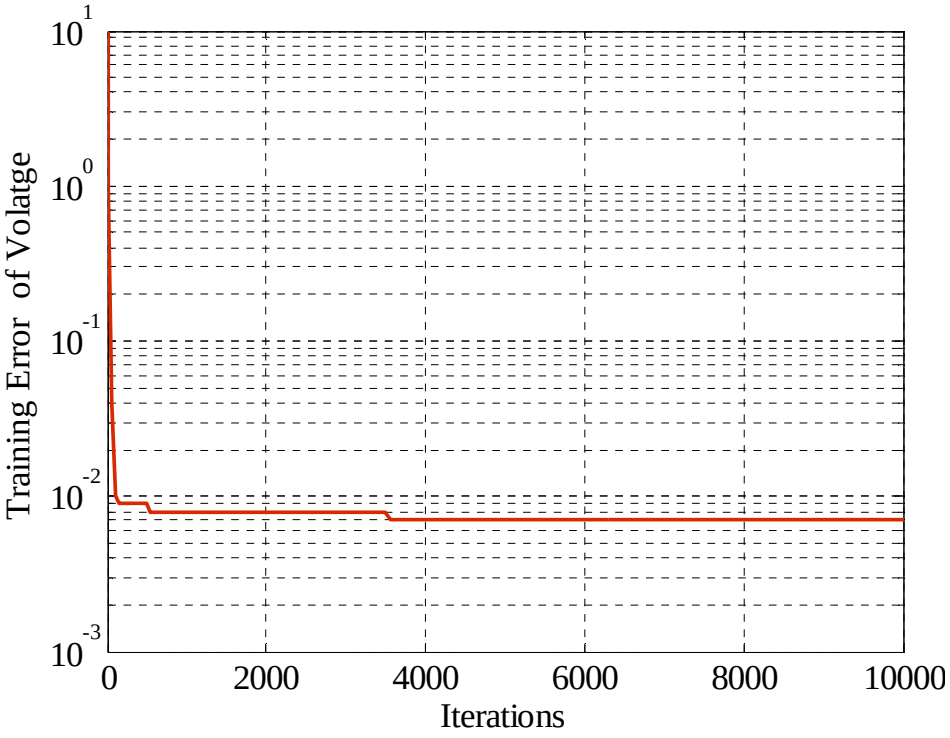


Figure 3.11. ANN's training Error of Voltage.

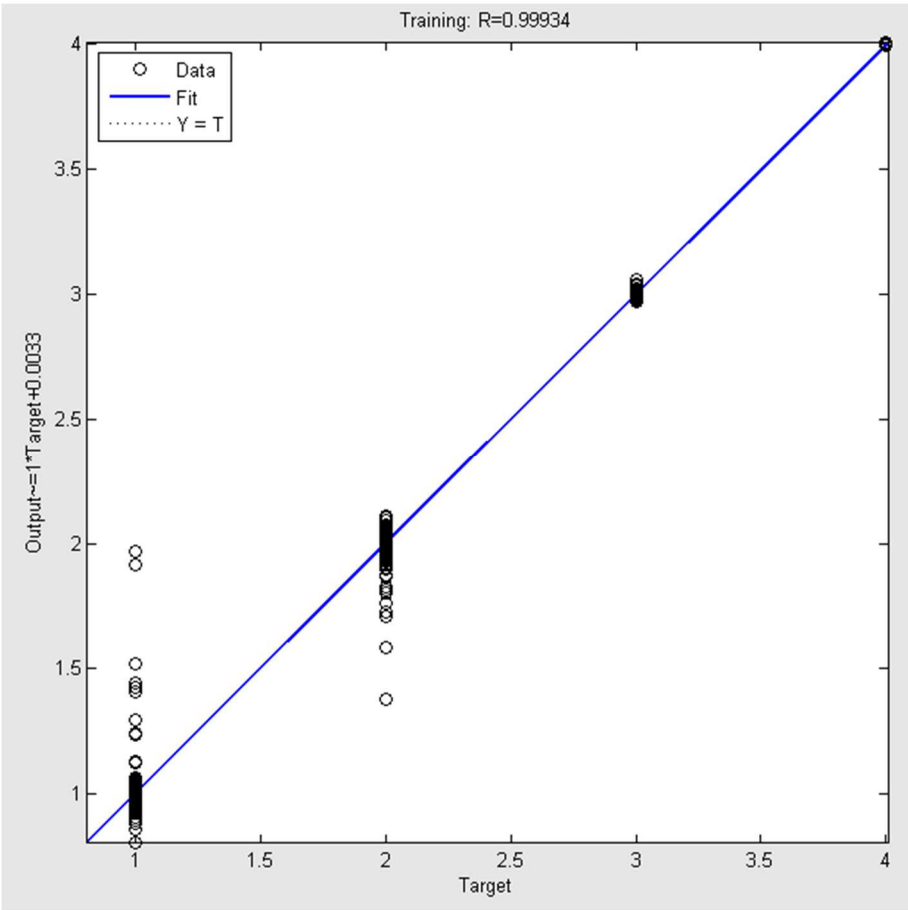


Figure 3.12. Classification accuracy of Voltage.

b- Validation phase

The classification results are illustrated in Figures 3.13 the second ANN, which shows that the ability of the obtained ANNs to introduce the incoming data in their right classes with high accuracy displaying 99% for the voltage classification. Some confusion classes are between the first class and the second as well as between the second and the third class, this confusion is due to the temperature variation.

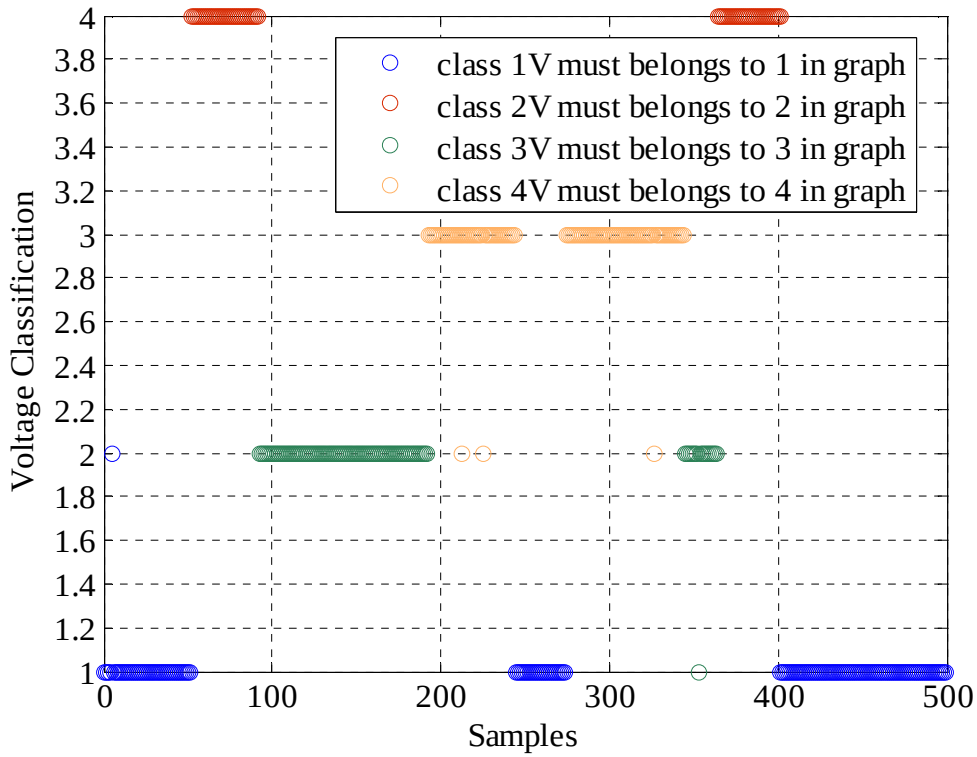


Figure 3.13. Vmpp classification using ANN.

3.5.1.3. Global classification of faults in PV generator.

As illustrated by Figure 3.14, the majority of the output decisions are in their right classes, the few confusion cases are 1) between healthy system (C1) and one PV module short circuited (C2), and between two PV modules short circuited (C3) and one PV module short circuited (C2) and 2) two confusion cases for healthy system (C1) belonging to faulty string (C5). The analysis of these confusion cases revealed that their causes are the variation of the irradiance and the temperature respectively. However, notwithstanding these isolated cases of confusion, the final simulation results show a high performance with a good accuracy equal to 98.6%.

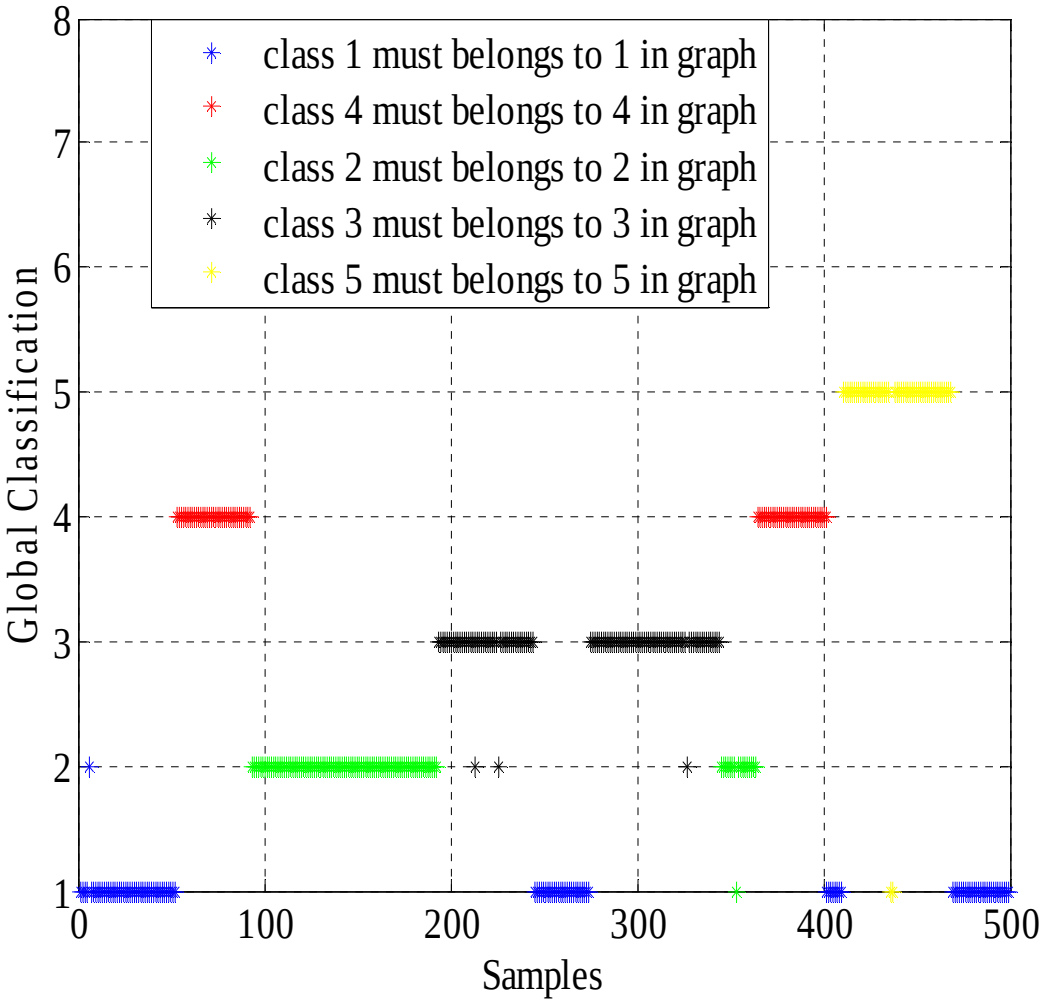


Figure 3.14. Global diagnosis of the system.

Table 3.6. Faults correspondence table for ANN test phase.

Algorithm	Levenberg-Marquardt	Levenberg-Marquardt
Hidden layer	2	1
Number of neuron in each hidden layer	8 x 8	45
Iterations	2800	1750
Performance	0.0001	0.001
Time (min)	7.52	6.43
R (Coefficient of re-gression)	0.9808	0.99934

Table 3.7. Device specifications of learning machine.

Device name	Ip 157
Processor	Intel(R) Core(TM) i5-4310U CPU@2.00GHz 2.60GHz
Install RAM	8.00GB
Device ID	E07ADB3E-6E21-453A-B746-28B1A825C51A
Product ID	00330-80000-00000-AA288
System type	32-bit operating system.x32-based processor

Table 3.6 resumes the values of criterions existing in ANNs algorithms, which plays a significant role in classification of both electrical parameters. In addition, Table 3.7 demonstrates the device specifications of learning machine used in this approach.

3.5.2. Global test Classification

To show the effectiveness of the developed IFD, a test considering new samples for each attribute has been performed in order to evaluate the obtained results. For this aim 120 samples are applied for each attribute and split into seventeen cases as summarized by Table 3.7. The results of this test are illustrated by Figure 3.15. The robustness of the IFD approach is evident since only two faulty situations among the 120 cases have been wrongly classified. The first case is due to the confusion between healthy system and one PV module short circuited. This confusion is caused by the temperature variation, knowing that if the temperature increases the voltage decreases and vice versa. The second is the result of confusion between the healthy system and the disconnected string.

Table 3.7. Faults correspondence table for ANN test phase.

Faults	Number of samples for each attribute
Healthy system	10 samples for each attribute.
One PV module short circuited	8 samples for each attribute.
Healthy system	4 samples for each attribute.
Disconnected string	6 samples for each attribute.
Healthy voltage	6 samples for each attribute.
Two PV module short circuited	8 samples for each attribute.
Healthy voltage	6 samples for each attribute.
Four PV module short circuited	11samples for each attribute.
Healthy system	6samples for each attribute.
Disconnected string	5 samples for each attribute.
Two PV module short circuited	8 samples for each attribute.
Healthy system	7samples for each attribute.
Four PV module short circuited	7samples for each attribute.
One PV module short circuited	4 samples for each attribute.
Healthy system	4 samples for each attribute.
Disconnected string	11 samples for each attribute.
Healthy system	9 samples for each attribute.

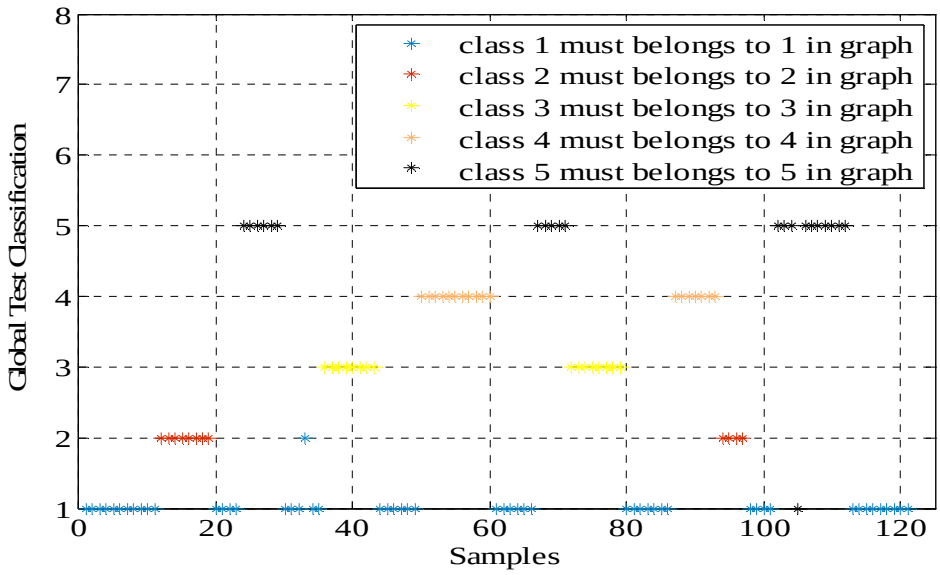


Figure 3.15. IFD approach test results.

This confusion is due to the irradiance variation, knowing that the irradiance is proportional to the current in increase or decrease. The accuracy of the IFD approach for this test is equals to 98.34%.

3.5.3. Performance of the system

The performance's evaluation of the proposed algorithm is mandatory. In fact, for each fault, the total samples are used for the test of evaluation. In this case, 524 samples are used for training phase and 500 samples for the classification phase for each attribute in two ANNs. The table below Table 3.8 demonstrates the developed algorithm's decision for each class in percentage. The best obtained results of the performance of the system are colored in red on the diagonal with mean global precision displaying 98.6% that represents the overall accuracy. This percentage signifies an excellent result of localization and classification of fault diagnosis in PV generator.

Table 3.8. Performance of classification with ANN (%)

		Decision Class				
		C 1	**C 2**	**C 3**	**C 4**	**C 5**
Input Class	**C 1**	99.18	0.82	0	0	0
	C 2	0.83	99.16	0	0	0
	C 3	0	4.77	95.23	0	0
	C 4	0	0	0	100	0
	C 5	0	0	0	0	99.6

Overall accuracy=98.6%.

3.6. Conclusion

The fault detection and diagnosis of PV systems is necessary not only to increase system power generation reliability but also for operating costs reduction. This approach proceeds to elaborate an important database for healthy and faulty operation, thereafter, to classify the faults information using two trained ANNs and finally to identification and recognize the corresponding fault.

In this chapter a new intelligent algorithm for PV systems' diagnosis and fault detection (IFD) for grid-connected photovoltaic systems is presented. It guarantees four faulty operating cases: one PV module short circuited in PV string, two PV modules short circuited in PV string, four PV modules short circuited in PV string and one string modules disconnection in a PV array. For a high efficiency of the diagnosis each electrical parameter, which is considered as the fault signature, is classified separately. The final diagnostic tests were successful with matching rates equal to 99.6% and 99% for the classification stage and 98.6% for the fault identification and isolation during the final phase of the diagnosis process. After obtaining excellent results in diagnostic, another question has arisen; what is the impact of the choice of ANNs on this diagnosis approach? That is discussed in the next chapter.

CHAPTER 4 COMPARATIVE STUDY OF PV GENERATOR DIAGNOSIS WITH FOUR DIFFERENT NEURAL NETWORKS.

4.1. Introduction

In response to the question asked in the previous chapter, a comparative study of diagnosis in PV generator has been proposed in this chapter in order to analyze the impact of the ANN's choice on the diagnosis quality. For this reason, four types of Artificial Neural Network have been proposed with five different algorithms that are substituted in the same IFD algorithms, and their performances are analyzed and compared to provide a well-argued response to the previous question. The five neural networks have the same four inputs: solar irradiation, cell temperature, the current and voltage of the maximum power point of the I-V characteristic corresponding to the first two inputs (the working conditions).

4.2. Methodology

Presently, fault diagnosis becomes the modern subject in PV installations that takes an important place in the world in order to guarantee their safety and reliability. For this reason, the accuracy, the sensivity, the specificity and the precision of fault detection and isolation are the most pertinent criterions of the diagnosis quality. This chapter provides analysis of the impact of the Artificial Neural Network choice on these criterions. To achieve this goal, five ANNs are studied: Back-Propagation Neural Network (BPNN), Probabilistic Neural Network (PNN), Generalized Regression Neural Network (GRNN) and two Radial Basis Function Neural Network (RBF1, RBF2). These types of ANNs are used to identify and locate the most frequently fault encountered in PV installations such as: open circuit and short circuit fault in PV generator. The comparison study used the same PV installation, the same working conditions, the same data and the same types of faults with the five algorithms.

4.3. Comparative study

This comparative study is structured in two phases. A first phase aimed at choosing the ANNs to be tested while the second part concerns their effectiveness evaluation.

4.3.1. Phase 1: ANNs choice

ANNs have been used with succeed in classification, pattern recognition, fault detection and diagnosis. To study the impact of the ANNs type of the performance of the IFD algorithm, several ANNs have been considered by maintaining the whole algorithm topology and by using the same learning and testing data and conditions. The five ANNs chosen for this comparative study are 1) Back Propagation Neural Network (BPNN), 2) statistical ANNs with two Radial Basis Function Networks which are considered to be the most important statistical neural networks in the literature. In the present study, two RBF ANNs are used and noted RBF1 and RBF2. The fourth ANN is a Probabilistic neural network (PNN) and the fifth ANN is a Generalized Regression Neural network (GRNN) both belonging to the previous family of statistical ANNs.

4.3.1.1. BPNN

The origin of a back propagation neural network (BPNN) is a multilayer perceptron (MLP) that contains three basic parts shown in Figure 4.1, the first part represents the input layer responsible to receive information, the third part represents the output layer captures and centralizes process information, while in the second part situated between the two last parts represents the hidden layer, which is composed of at least two layers of nodes [65, 66].

The last layer cited above signifies a particular importance in neural network, which implements the nonlinear transformations to the inputs entry into the network. Furthermore, the hidden layers are the mathematical functions applied weights to the input data adding bias $v = \sum_{i=0}^{n} x_i w_i + b_i$ and conduct them via an activation function as the sigmoid function shown in Equation 4.1 in sequence to bring an output particular to a considered result.

$$\varphi_i(x) = \frac{1}{1+\exp(-v)} \qquad (4.1)$$

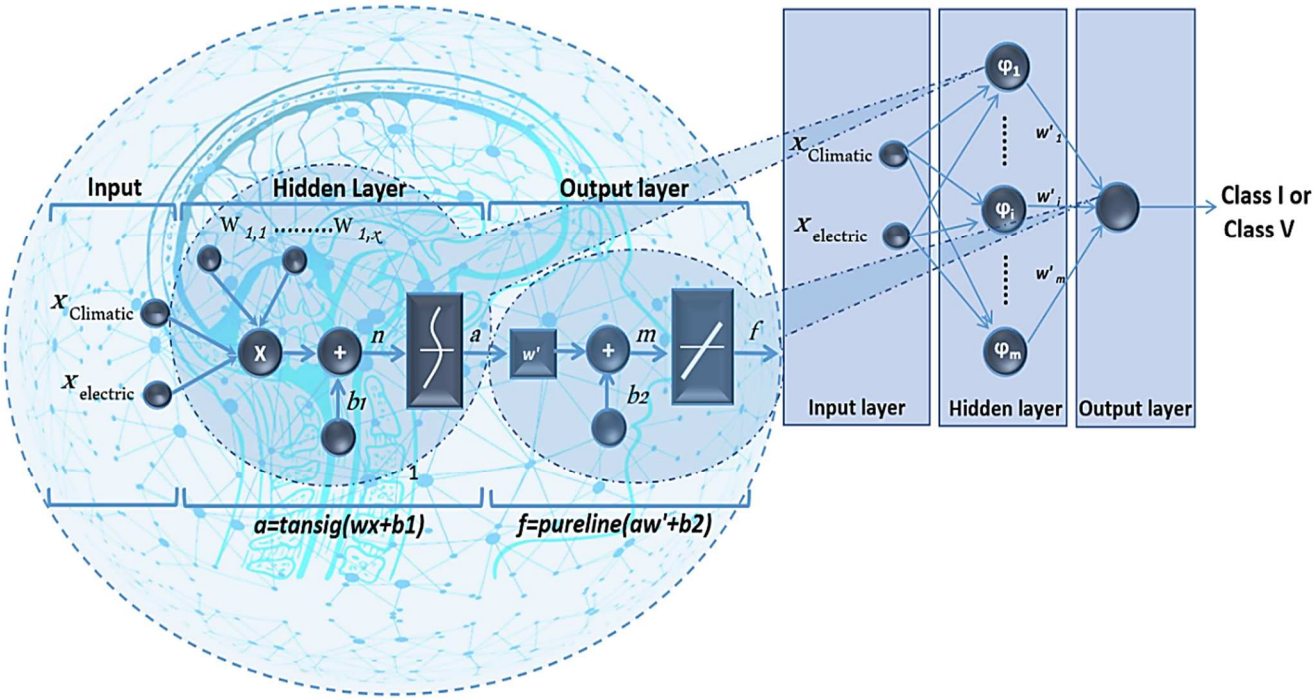


Figure 4.1. Architecture of BPNN model of the present work.

In other words, the term Back propagation neural network (BPNN) applied on MLP is used for the reason that it contributes to adjusting the weights of the neurons with the objective to achieve a results more and more close to the real result [51].

4.3.1.2. Statistical neural network

This type of neural network uses statistical methods and probability theory to compare a number of random variables as probability of density function (PDF) aiming to obtain the exact decision[67].

a.Radial Basis Function Network (RBF)

The radial basis function network is dissimilar to classical multilayer perceptron neural network. The particularity of radial basis function is to be the central point of radial basis function neural network and to be advantageous in classification mode for nonlinear data, it contains a simple structure and its convergence is speed and fast [68]. As classical neural network, three layers compose RBF network, each layer comprises a specific function different to other layer. Figure 4.2 corresponds to the current neural network as mentioned by its proper function. In more depth, the first layer represents the input layer where an equivalent number between the nodes and the dimensions of the input is announced, the third layer characterizes the output layer knowing that the number of nodes is similar to the size of output data, The particularity of this layer is mapping the nonlinearity as a linear combiner involved in a novel space. In the middle, the hidden layer called intermediate layer is introduced, noting that this layer is nonlinear and each node in this layer is deliberately identified and characterized through an activation function φ [69]. The action between input and hidden layer is unsupervised contrary to the action between hidden and output layer that is supervised.

The radial basis function process estimates each data of the input vectors, the recoded training data collected from the present network are compared to the input value in order to produce a similar data, each similarity value is multiplied by weights in hidden layer and finally the summation is displayed in the output layer. For any new input data, the network can simply be calculated through to a Euclidean distance measurement between the input and training data. In this kind of cases, it is necessary to determine the center and the standard deviation or spread σ of node coming from the intermediate layer as well as the weight matrix disposed between the hidden layer and the output layer. In general, a possibility to determine the center of node in hidden layer can be ensured by cluster method. This method requires K-Means Clustering Algorithm in sequence to divide data points into diverse categories, the favored center point is more particular when a similarity of characteristics and properties exists in the same type of internal. The activation function is radially symmetric basis function presented as Gaussian function in Equation 4.2 [70, 71]:

$$\varphi_i(x) = exp\left(-\frac{\|x-c_i\|}{\sigma_i^2}\right) \quad (4.2)$$

Where *x* represents the meteorological parameters; c_i and σ_i are center and spread of the i[th] RBF node, respectively. The spread can be calculated by Equation 4.3 [68]:

$$\sigma_1 = \frac{d_{max}}{\sqrt{2n}} \quad (4.3)$$

Where n is the number of the node in the intermediate layer, d_{max} is the maximum distance between the cluster centers selected

Then, the outputs of the non-linear activation ($\varphi_i(x)$) are integrated linearly with the weight vector ω_i of the output layer to produce the output network *class m* Equation 4.4:

$$class\ m = \sum_{i=0}^{n} \varphi_i w_i \quad (4.4)$$

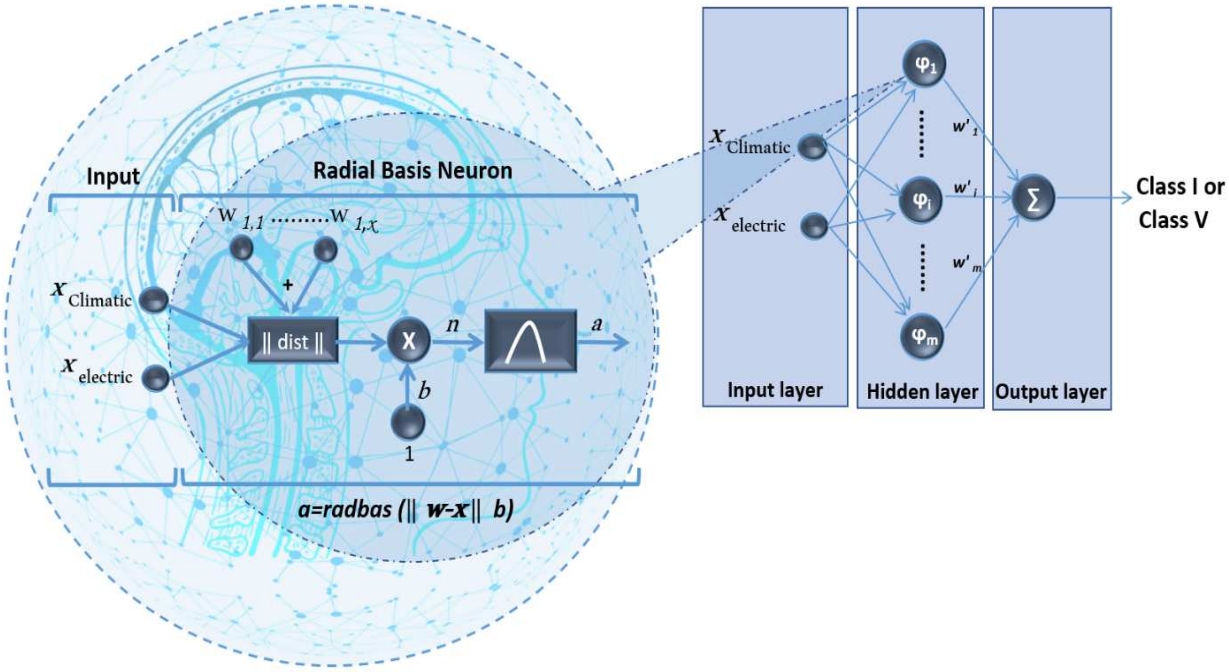


Figure 4.2. Architecture of RBF model of the present work.

Two methods of radial basis function neural network with two functions are presented below:

- **RBF1**

This type of neural network is presented by *newrbe* function. It is very quickly concepting a radial basis function creating a network with zero error on training vector. Spread should not be so large that each neuron is effectively responding in the same, large, area of the input space [72].

- **RBF2**

This type of neural network is presented by *newrb* function. It iteratively creates a radial basis network one neuron at a time. The larger spread is the smoother the function approximation. Too large a spread means a lot of neurons are required to fit a fast-changing function. Too small a spread means many neurons are required to fit a smooth function, and the network might not generalize well. The Call of *newrb* with different spreads is necessary to find the best value for a given problem [72].

b. PNN

According to the literature, the probabilistic neural network (PNN) designed by Specht [58] is considered as favor supervised learning network in reason to its short period of time training ability, perfect generalization qualification, excellent particularity in pattern recognition as good as classification and diagnosis [51, 64, 79]. About its architecture, the PNN contains four layers: the input layer responsible of the data entry, the pattern layer destined for calculating the probabilities of each class, the summation layer destined to sum the probabilities obtained from the last precedent layer multiplied by their respective weights coefficient given and finally the output layer displays the final classification where the input data belongs [73,74].

The pattern unit is responsible to receive information from the input units x_i and corresponded it to their respective weights coefficient given W_i [51]. In fact, the PNN performance is guaranteed by two predominant factors, the first one is the number of neurons in pattern layer and the second one is suitable activation function given by the Equation 4.5.

$$\varphi(x) = \exp\left(-\frac{(w_i - x)^t (w_i - x)}{(2{\sigma_1}^2)}\right) \quad (4.5)$$

Where: σ represents the smoothing parameters (spread)

Appropriately, smoothing parameter plays an important role during the PNN model optimization process, it depends to the input data. The output units of the pattern layer will be transfer to the units of summation layer, in this part a certain number of units are existent referring to the number of classes it means one unit per class. Each output unit in the summation layer calculates probability density function (PDF) of the input vector given in Equation 4.6 [51, 73, 74]:

$$f_a(x) = \frac{1}{(2\pi{\sigma_1}^2)^{M/2}} \frac{1}{n} \sum_{i=1}^{n} exp\left[-\frac{(w_i - x_{ai})^t (w_i - x_{ai})}{2{\sigma_1}^2}\right] \quad (4.6)$$

Where: *M* is the number of patterns, n is learning set size, x_{ai} describes the corresponding i_{th} training pattern of a class. The output layer called decision layer contains one unit, which decides and displays the final class coming from summation layer, which can be defined as:

$$class(x) = argmax\{f_a(x)\}, \qquad 1 \le i \le M \quad (4.7)$$

Where: *M* is number of classes, *class(x)*: denotes the predicted class of *x*.

Figure 4.3 illustrates the architecture of this kind of neural network plus their functions according to exact layer.

Ensuring the effectiveness of PNNs is attached by the PDF accuracy depending of the excellent determination value of smoothing parameter σ_I, which symbolizes a huge advantage for this type of network.

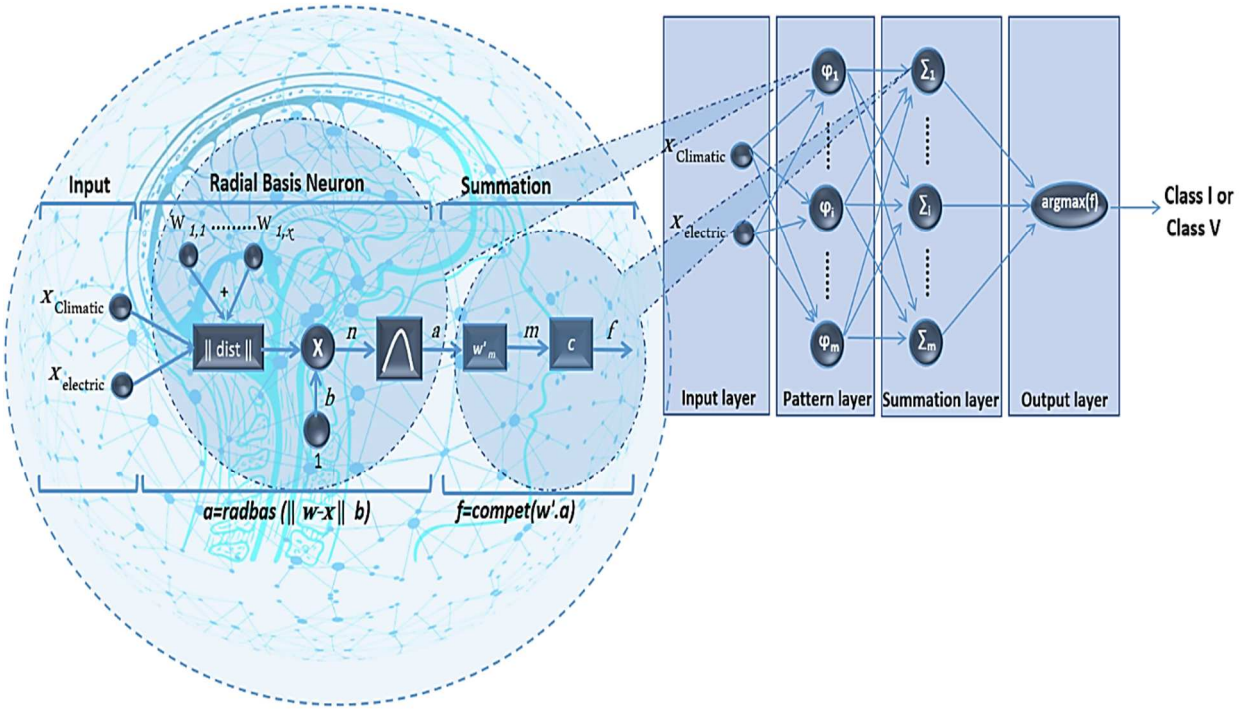


Figure 4.3. Architecture of PNN model of the present work.

c. GRNN

As probabilistic neural network (PNN) generalized regression neural networks (GRNN) are recognized by their quick capability in training phase on spare data set [60]. Its architecture contains four layers: the input layer, the pattern layer, the summation layer and the output layer as shown in Figure 4.4. The role of input layer is just like others ANNs responsible to receive the information data. When in the pattern layer plays the same role as pattern layer in PNN, its equation is as follows [74, 75]:

$$f_a(x,y) = \frac{1}{(2\pi{\sigma_1}^2)^{M/2}} \frac{1}{n} \sum_{i=1}^{n} \left\{ exp\left[-\frac{(X-X_i)^t (X-X_i)}{2{\sigma_1}^2} \right] exp\left[-\frac{Y-Y_i}{2{\sigma_1}^2} \right] \right\} \quad (4.8)$$

Where:

n is number of simple observations and M is dimension of the vector variable x, σ is smoothing parameter, X is particular measured value of the random variable x which represents the independent data in the input system, the regression performed by GRNN products the most probable scalar Y provided from specified input vector x,

which represents the dependent data in the output system, Yi is desired scalar output given the observed input Xi.

The summation layer has two kinds of processing units: the first one is called the “Numerator” (N) neuron representing the sum of the pattern layer outputs, in this case, each weighted by an observed output scalar, *Yi*, corresponding to *Xi* in the training samples, defined as follow Equation 4.9:

$$Numerator(N) = \sum_{i=1}^{n} Y_i \, exp\left(-\frac{(X-X_i)^t(X-X_i)}{2{\sigma_1}^2}\right) \qquad (4.9)$$

While the second one is called the “denominator” (D) neuron representing the sum of the pattern unit outputs, presented as follow Equation 4.10:

$$Denominator(D) = \sum_{i=1}^{n} exp\left(-\frac{(X-X_i)^t(X-X_i)}{2{\sigma_1}^2}\right) \qquad (4.10)$$

At the end, the output layer contains just one neuron displaying the classification , receives specifically the two outputs from the summation units and divides the “Numerator” part by the “Denominator” part Equation 4.11 to produce an estimate for *y* given *X*.

$$\hat{C}(x) = {Numerator}/{Denominator} \qquad (4.11)$$

$\hat{C}(x)$ means the final electrical parameter classification.

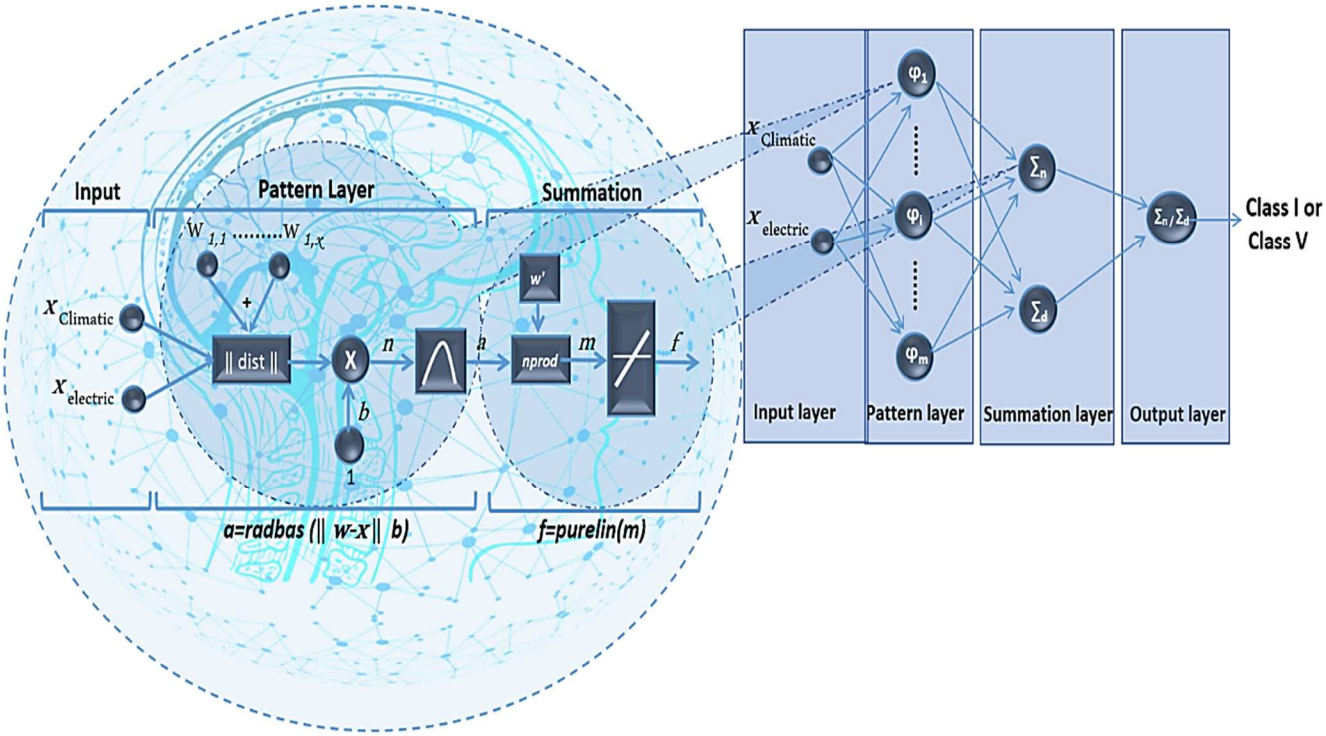


Figure 4.4. Architecture of GRNN model of the present work.

4.3.2. Phase 2: Effectiveness evaluation

In order to evaluate the effectiveness of the five IFD algorithms, theirs results are analyzed using the most frequently used diagnosis performance indicators in science and engineering fields [77, 78]:

a. **Accuracy:** implies how nearest is the results to the real value.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (4.12)$$

b. **Sensivity:** measures in what way the positive samples are correctly classified.

$$Sensivity = \frac{TP}{TP+FN} \times 100 \qquad (4.13)$$

c. **Specificity:** measures in what way the negative samples are correctly classified.

$$Specificity = \frac{TN}{FP+TN} \times 100 \quad (4.14)$$

d. **Precision:** implies how nearest the results are to each other.

$$Precision = \frac{TP}{FP+TP} \times 100 \quad (4.15)$$

Where:

TP: true positive, signifies that the samples contain characteristics of a specific class and indeed they are classified in this class.

TN: true negative, signifies that the samples do not contain characteristics of a specific class and indeed they are not classified in this class.

FP: false positive, signifies that the samples do not contain characteristics of a specific class and they are classified in this class.

FN: false negative, signifies that the samples contain characteristics of a specific class and indeed they are not classified in this class.

Table 4.1 summarizes the four major categories as result of binary classification containing two rows and two columns into confusion matrix called confusion table in the intension to confirm the performance evaluation related to the classifier. The number of rows and columns depends on the number of classes. The terms true and false refer to whether the prediction corresponds to the external criticism conversely to the terms positive and negative that refer to the prediction of the classifiers.

Table 4.1. Confusion matrix under intermittent classification troubles.

Classification outcome from experimental data / Classification outcome from ANNs		real label	
		True class	False class
predicted label	True class	TP	FP
	False class	FN	TN

4.4. Results presentation and discussions

This section depicts the central and the most important part of this dissertation, which consists to simulate the diagnosis of the model with five different algorithms. Each algorithm diagnoses the model separately from the other to find which one is the best in terms of response time, efficiency and accuracy. Therefore, two ANN's trained and tested for each algorithm, the outputs of both ANNs are inserted as the input of combinational algorithm in the interest to obtain the final classification.

The obtained database presents 12544 samples of each attribute, divided in two parts: 6272 samples for the ANN current classification and 6272 samples for the ANN voltage classification.

The first ANN classifies the current of the maximum power point. This ANN contains two neurons in input layer, which represent irradiance and current for the maximum power point respectively, one neuron in output layer, which represent current classification, between these two layers, there are two hidden layers of eight neurons in each one. In the training phase, each sample contains the current value at the MPP (I_{mpp} (A) and the irradiance level (W/m2) that is (2800 x 2 =5600 data) as input. Where, 50% of the samples represent healthy operation conditions and remaining 50% represent the disconnected string. The step of current ANN diagnosis needs 336 samples

of each attribute (Irradiance (W/m2) and I_{mpp} (A)) divided into thirty-one cases respectively for testing and ensuring the classification stability and efficiency. Table 4. 2 summarizes all treated cases with their number of samples of corresponding manner.

The second ANN classifies the voltage of the maximum power point. This ANN contains two neurons in input layer and one neuron in the output layer. The input neurons receive cell temperature and voltage of the maximum power point respectively when the output neuron gives the voltage classification. The ANN contains one hidden layer of forty-five neurons. For its training, 2800 samples have been used for each attribute (V_{mpp} (V), Temperature (°C)) it means (2800 x 2= 5600 data) as input, for training phase; 25% samples represent healthy voltage and the remaining 75% samples represent different faults divided into three equal categories which are: 25% for one PV module short circuited, 25% for two PV module short circuited and 25% for four PV module short circuited. For each case there are 700 samples for each attribute. On the other side, the ANN diagnosis employs 336 samples of each attribute (V_{mpp} (V), Temperature (°C)) means (336 x 2 = 672 data) distributed in thirty-one cases respectively resume in Table 4.3.

Table 4.2. Faults correspondence table of current.

Faults	Number of samples for each attribute
Healthy voltage	9 samples for each attribute.
Disconnected string	11 samples for each attribute.
Four PV module short circuited	15 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Two PV modules short circuited	30 samples for each attribute.
One PV module short circuited	16 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Disconnected string	9 samples for each attribute.
Two PV modules short circuited	11 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
One PV module short circuited	6 samples for each attribute.
Healthy voltage	10 samples for each attribute.
One PV module short circuited	8 samples for each attribute.
Two PV modules short circuited	7 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
Healthy voltage	4 samples for each attribute.
Disconnected string	13 samples for each attribute.
One PV module short circuited	11 samples for each attribute.
Two PV modules short circuited	14 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Disconnected string	13 samples for each attribute.
Healthy voltage	5 samples for each attribute.
One PV module short circuited	9 samples for each attribute.
Four PV modules short circuited	12 samples for each attribute.
Disconnected string	12 samples for each attribute.
Healthy voltage	21 samples for each attribute.
One PV module short circuited	13 samples for each attribute.
Four PV modules short circuited	15 samples for each attribute.
Two PV modules short circuited	10 samples for each attribute.
Healthy voltage	4 samples for each attribute.

Table 4.3. Faults correspondence table of voltage.

Faults	Number of samples for each attribute
Healthy voltage	9 samples for each attribute.
Disconnected string	11 samples for each attribute.
Four PV module short circuited	15 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Two PV modules short circuited	30 samples for each attribute.
One PV module short circuited	16 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Disconnected string	9 samples for each attribute.
Two PV modules short circuited	11 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
One PV module short circuited	6 samples for each attribute.
Healthy voltage	10 samples for each attribute.
One PV module short circuited	8 samples for each attribute.
Two PV modules short circuited	7 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
Healthy voltage	4 samples for each attribute.
Disconnected string	13 samples for each attribute.
One PV module short circuited	11 samples for each attribute.
Two PV modules short circuited	14 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Disconnected string	13 samples for each attribute.
Healthy voltage	5 samples for each attribute.
One PV module short circuited	9 samples for each attribute.
Four PV modules short circuited	12 samples for each attribute.
Disconnected string	12 samples for each attribute.
Healthy voltage	21 samples for each attribute.
One PV module short circuited	13 samples for each attribute.
Four PV modules short circuited	15 samples for each attribute.
Two PV modules short circuited	10 samples for each attribute.
Healthy voltage	4 samples for each attribute.

The classification results of the five algorithms are illustrated as follow:

4.4.1. BPNN classification

The classification tests show that all samples for the classification of the I_{mpp} are in their right classes as illustrated by Figure.4.5. While nearly all samples for the classification of V_{mpp} are in their correct classes. Only one confusion case has been encountered during which, the ANN confused a healthy voltage with a short-circuited PV module as shown by Figure 4.6. This confusion is due to the temperature variation. As a result, the combination of the two ANNs classification results outputs reveals a very high accuracy. In fact, only one sample is confused (confusion between C1 and C2) on a dataset that contains 336 samples as illustrated by Figure. 4 7.

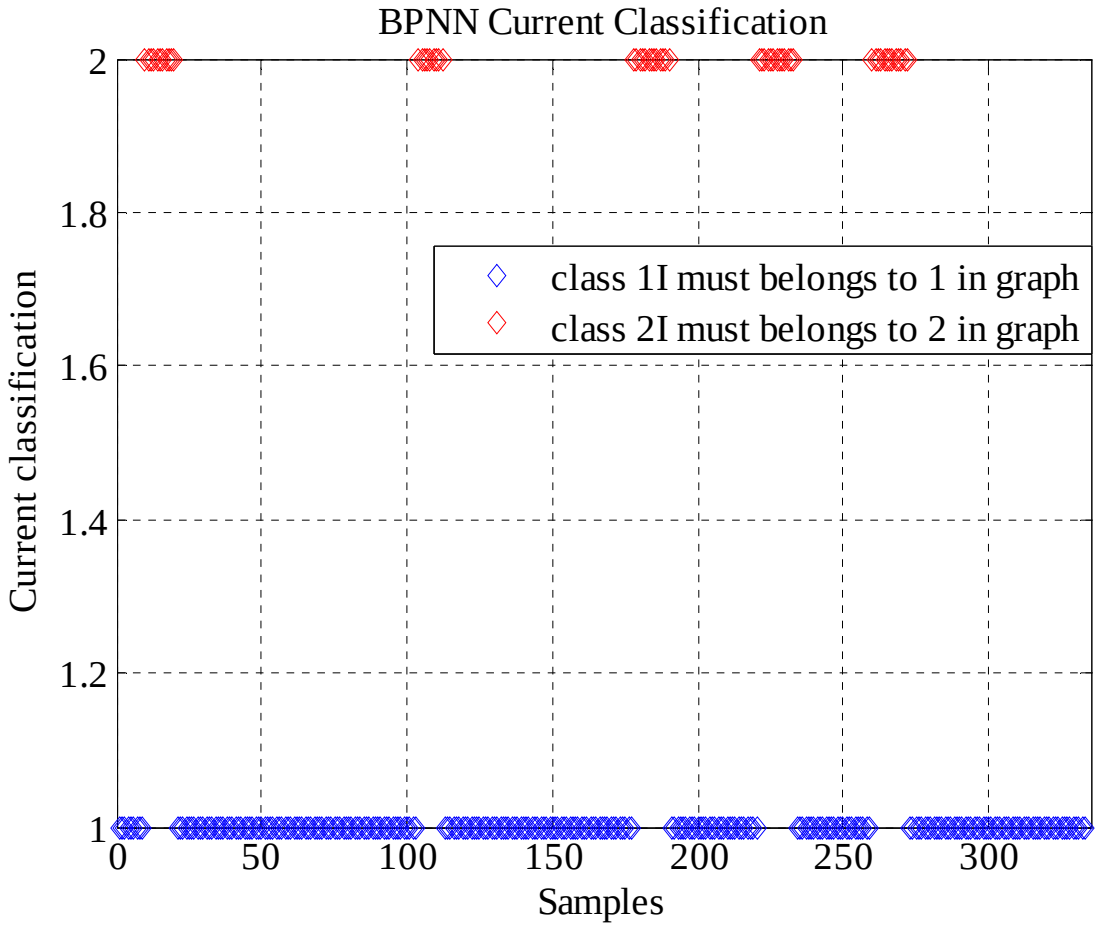


Figure 4.5. Impp classification using BPNN.

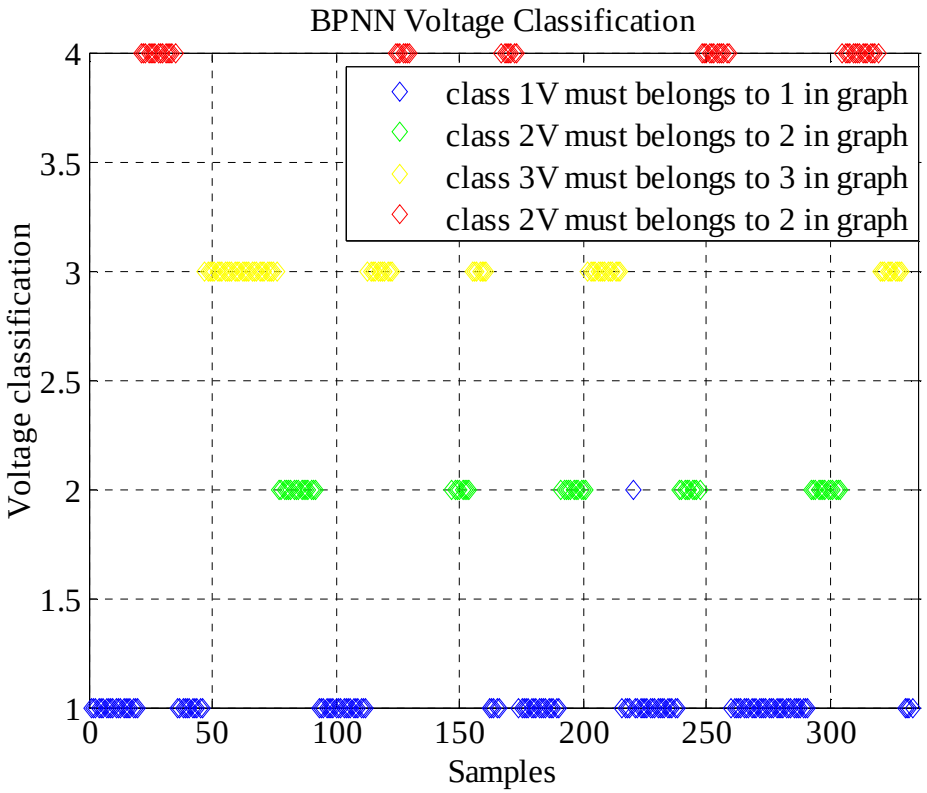


Figure 4.6. Vmpp classification using BPNN.

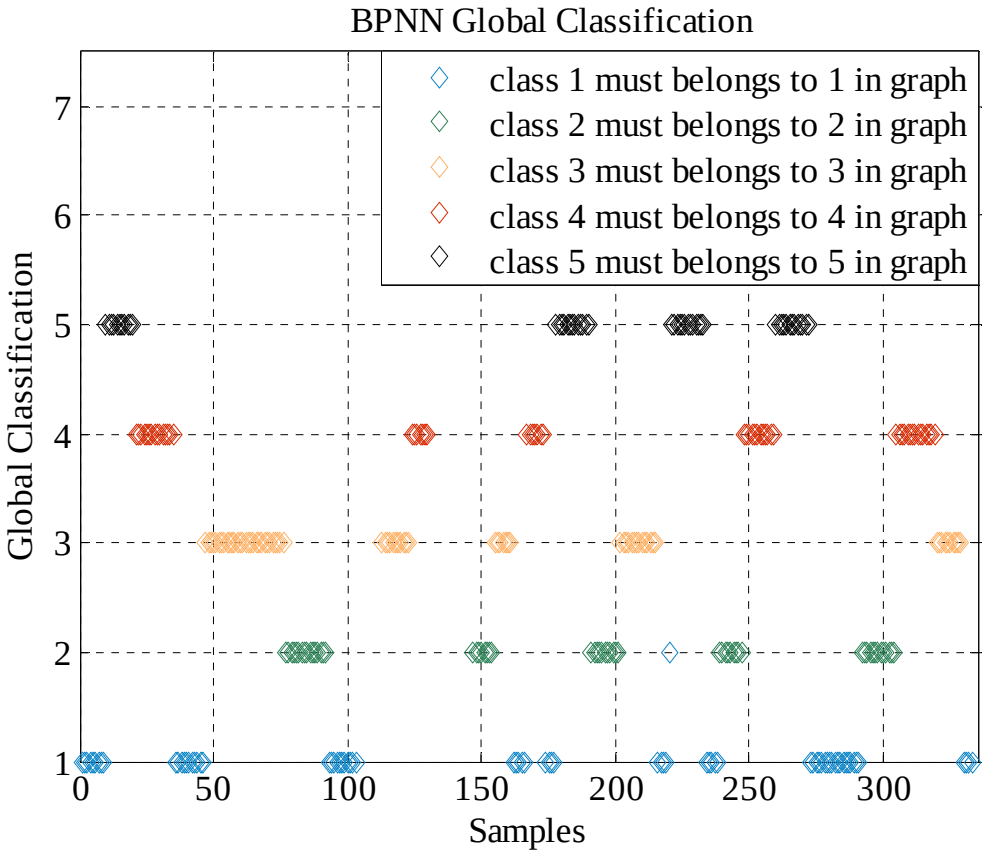


Figure 4.7. Global diagnosis of the system using BPNN.

4.4.2. RBF classification

4.4.2.1. RBF1

The classification results for this type of neural network are demonstrated in Figure 4.8 and Figure 4.9 representing the current classification and the voltage classification respectively. All samples for the classification of I_{mpp} by RBF1 are in their right classes, on one side. On the other side, nearly all samples for the classification of V_{mpp} by RBF1 are in their right classes only one confusion between one PV module short circuited and healthy voltage , the analyze of this case of confusion is due to the variation the temperature.

The combination of the data of both figures through this neural network is demonstrated in Figure 4.10 where the outcome reveals a very high accuracy, only one sample confuses between (C1) and (C2) on dataset of 336 samples.

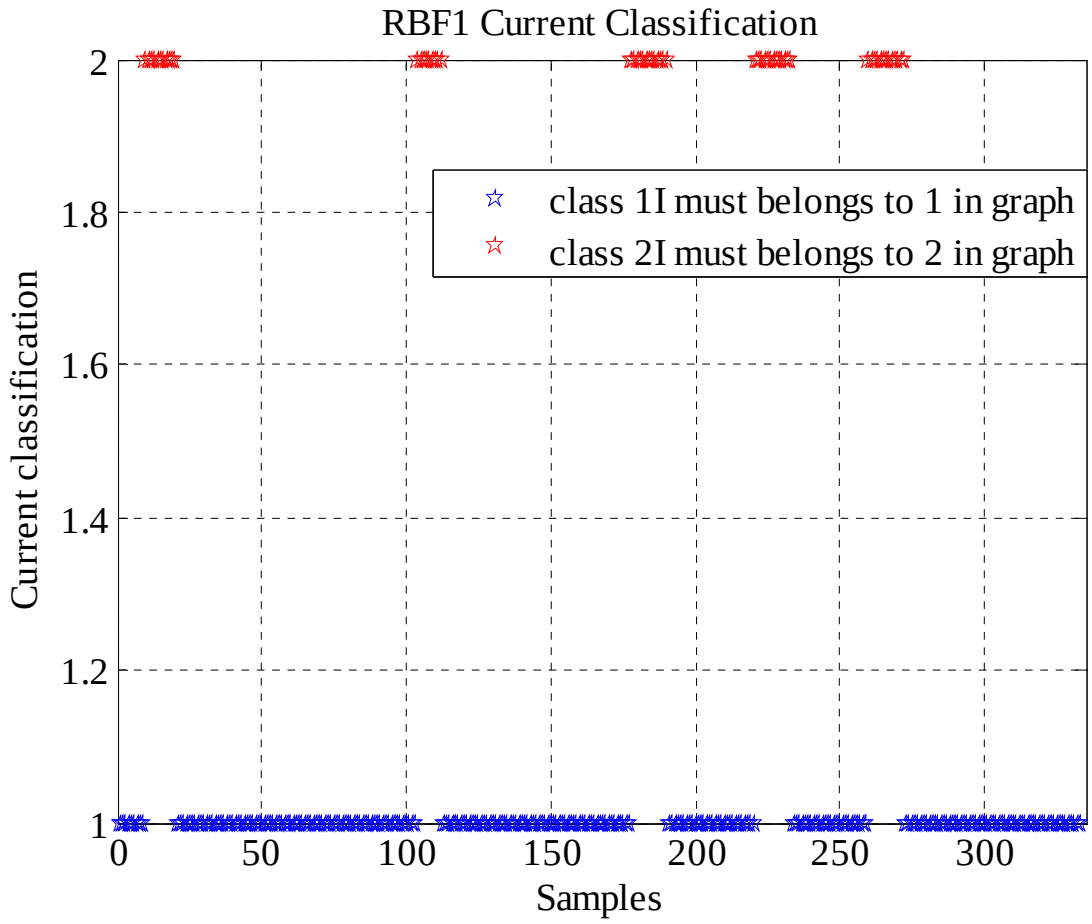


Figure 4.8. Impp classification using RBF1.

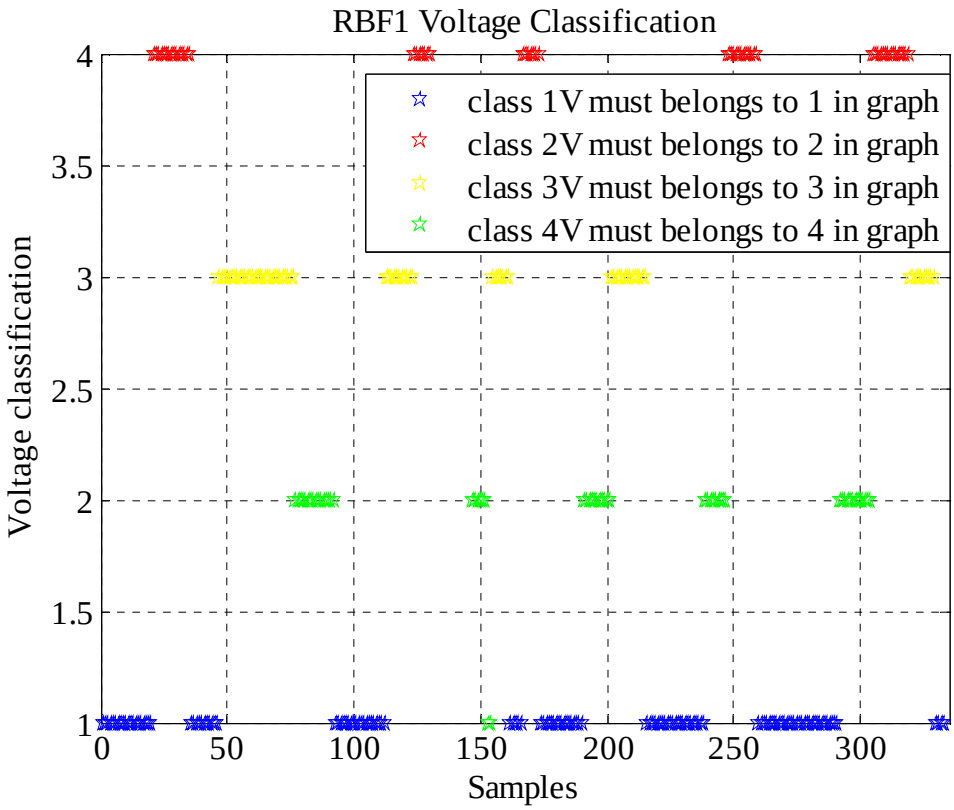


Figure 4.9. Vmpp classification using RBF1.

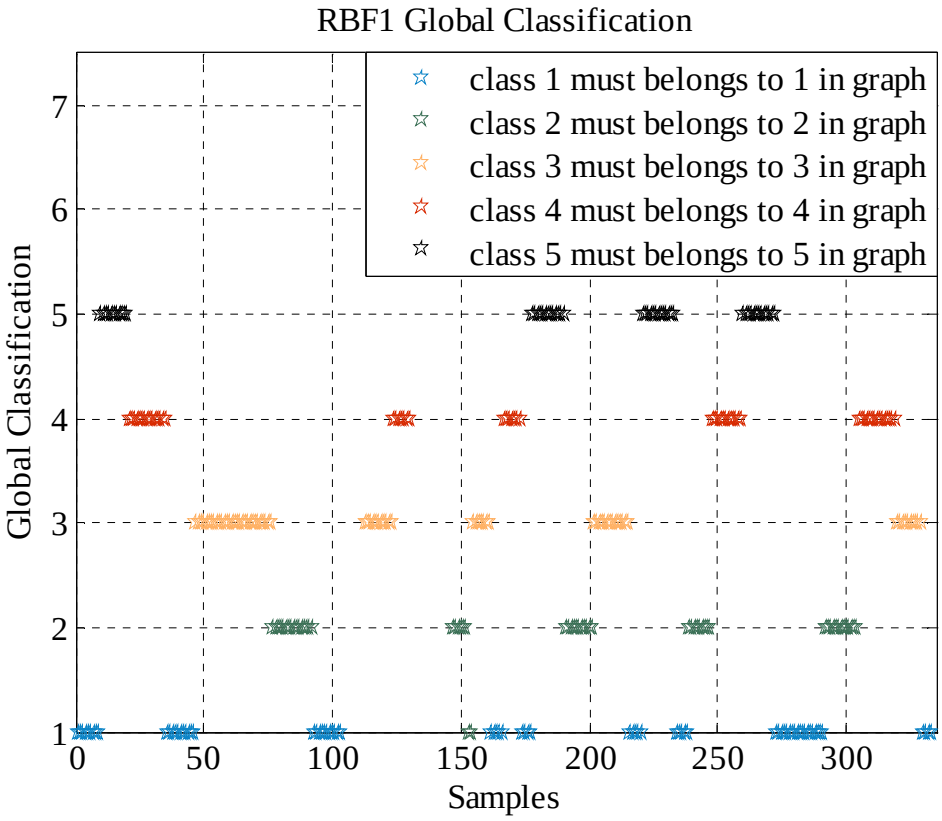


Figure 4.10. Global diagnosis of the system using RBF1.

4.4.2.2. RBF2

The classification results for the second type of neural network are demonstrated in Figure 4.11 and Figure 4.12 representing the current and the voltage classification respectively. All samples for the classification of I_{mpp} by RBF2 are in their right classes, on one side. On the other side, most samples for the classification of V_{mpp} by RBF2 are in their right classes but three confusion samples are between one PV module short circuited and healthy voltage and between two PV modules short circuited and one PV module short circuit , the analyze of this case of confusion is due to the variation the temperature.

The combination of the data of both figures through this neural network is demonstrated in Figure 4.13 where the outcome reveals a good accuracy, only one sample confuses between (C1) and (C2) on dataset of 336 samples.

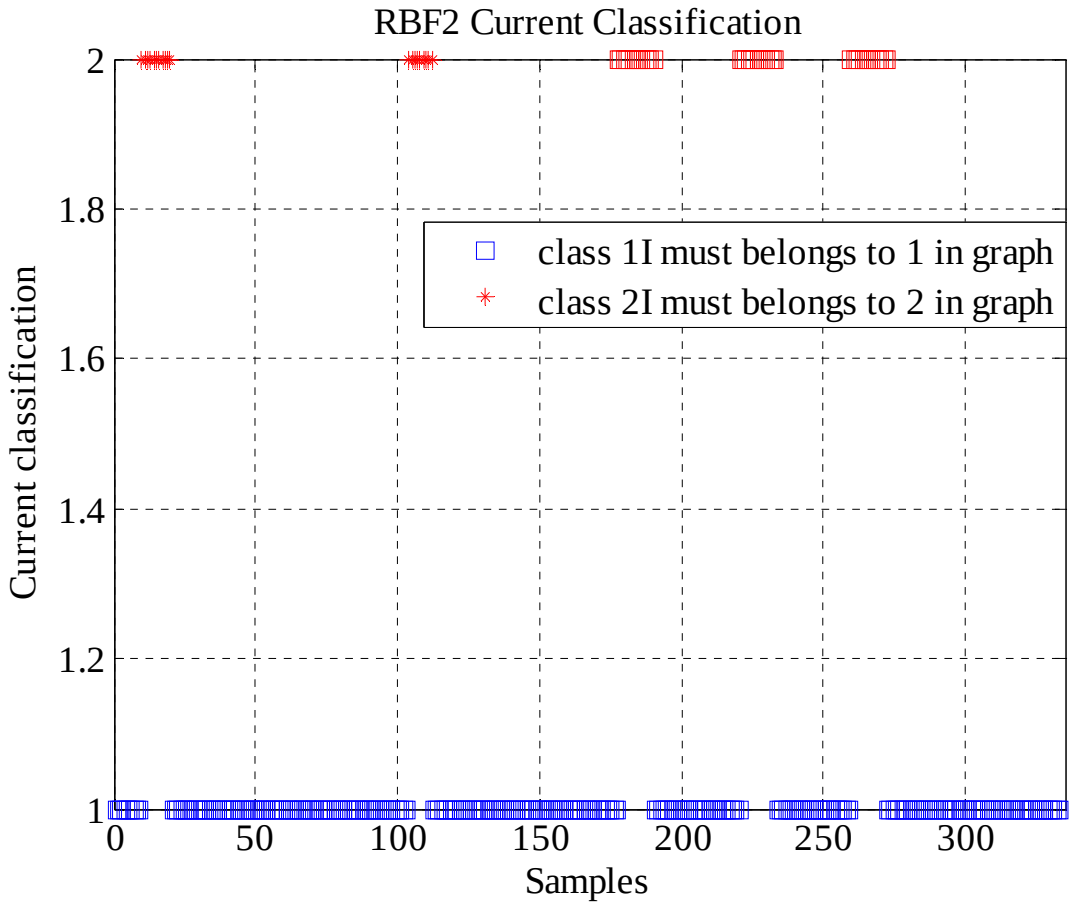


Figure 4.11. Impp classification using RBF2.

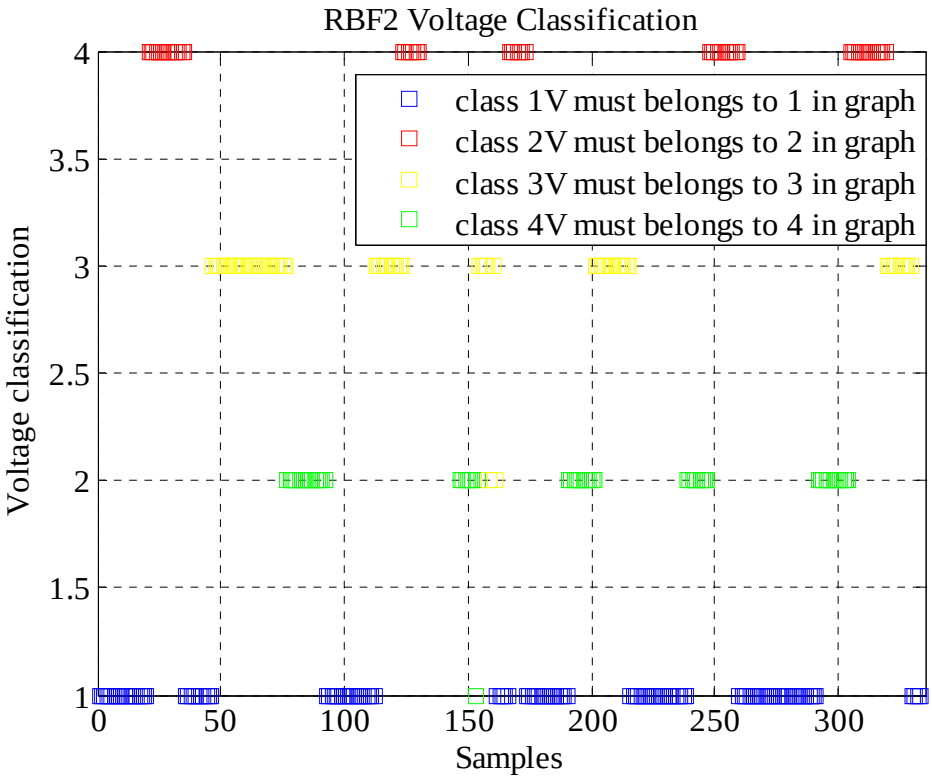


Figure 4.12. Vmpp classification using RBF2.

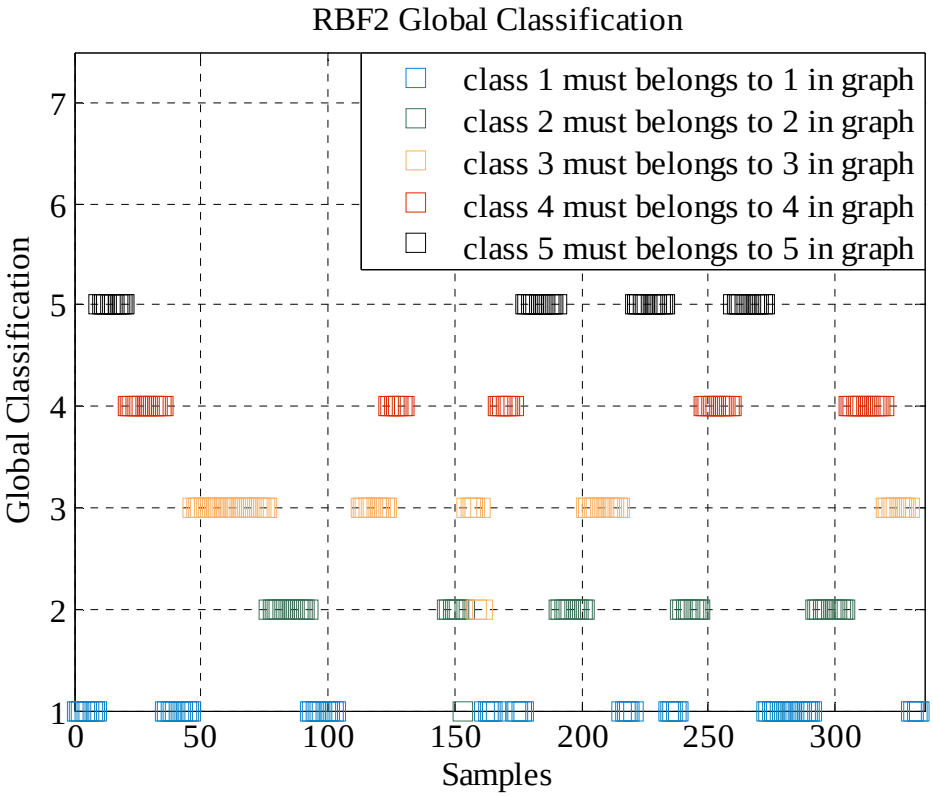


Figure 4.13. Global diagnosis of the system using RBF2.

4.4.3. PNN classification

For this model of neural network, the current and voltage classification are demonstrated in Figure 4.14 and Figure 4.15 respectively. All samples of the classification whether for the current or for the voltage are in their right classes without any confusion. The global diagnosis through this model of neural network is demonstrated in Figure 4.16 where the outcome reveals an excellent efficiency.

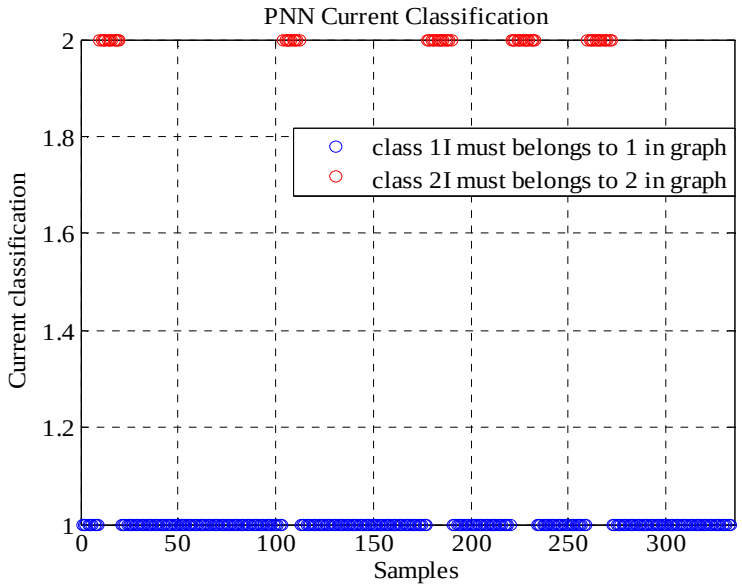


Figure 4.14. Impp classification using PNN.

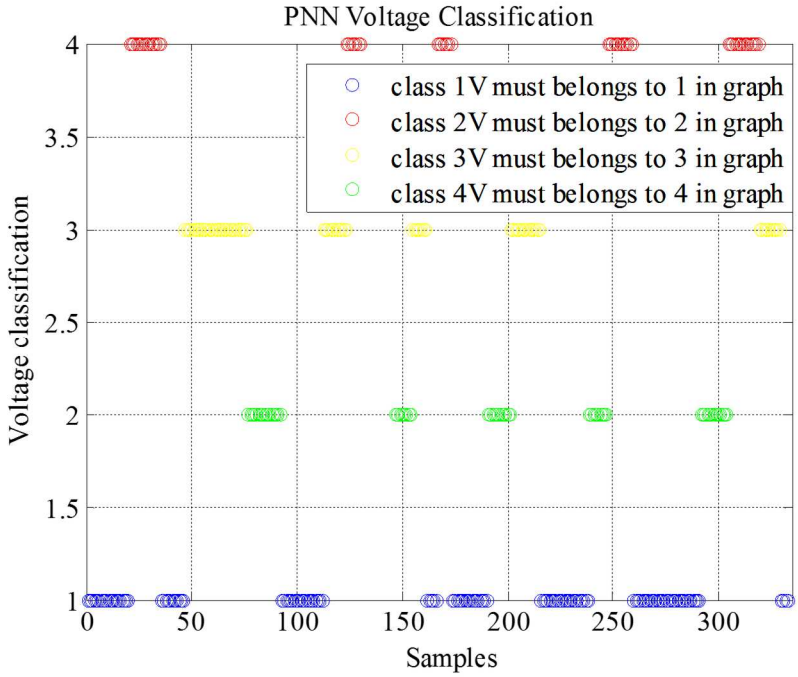


Figure 4.15. Vmpp classification using PNN.

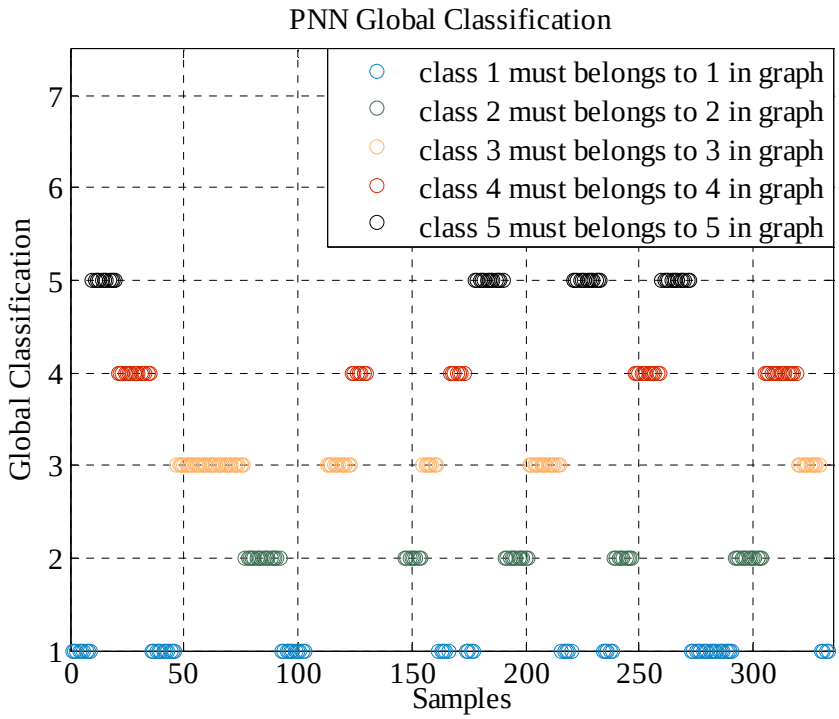


Figure 4.16. Global diagnosis of the system using PNN.

4.4.4. GRNN classification

The current and voltage classification for the last neural network are demonstrated in Figure 4.17 and Figure 4.18 respectively. All samples for the classification of I_{mpp} by GRNN are in their right classes and most samples for the classification of V_{mpp} by GRNN are in their right classes only two confusion samples are between two PV modules short circuited and one PV module short circuit , the analyze of this case of confusion is due to the variation the temperature.

The combination of the data of both figures through this neural network is demonstrated in Figure 4.19 where the outcome reveals a high efficiency, with two sample confused between (C2) and (C3) on dataset of 336 samples.

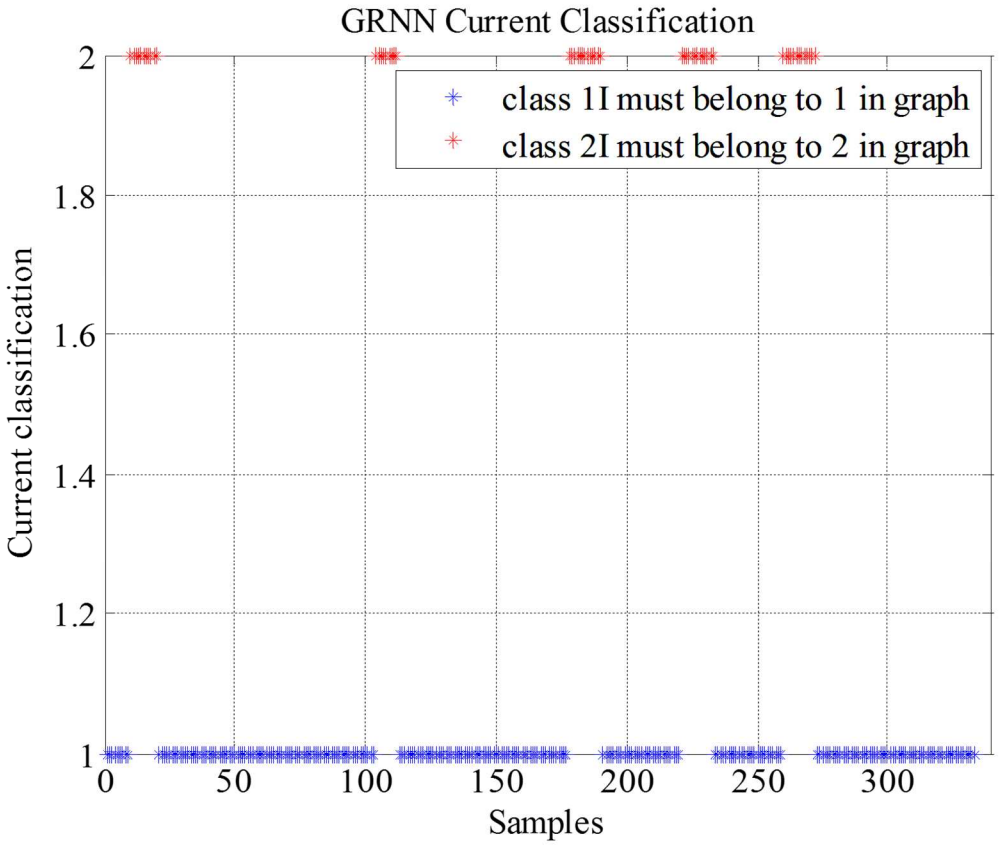


Figure 4.17. Impp classification using GRNN.

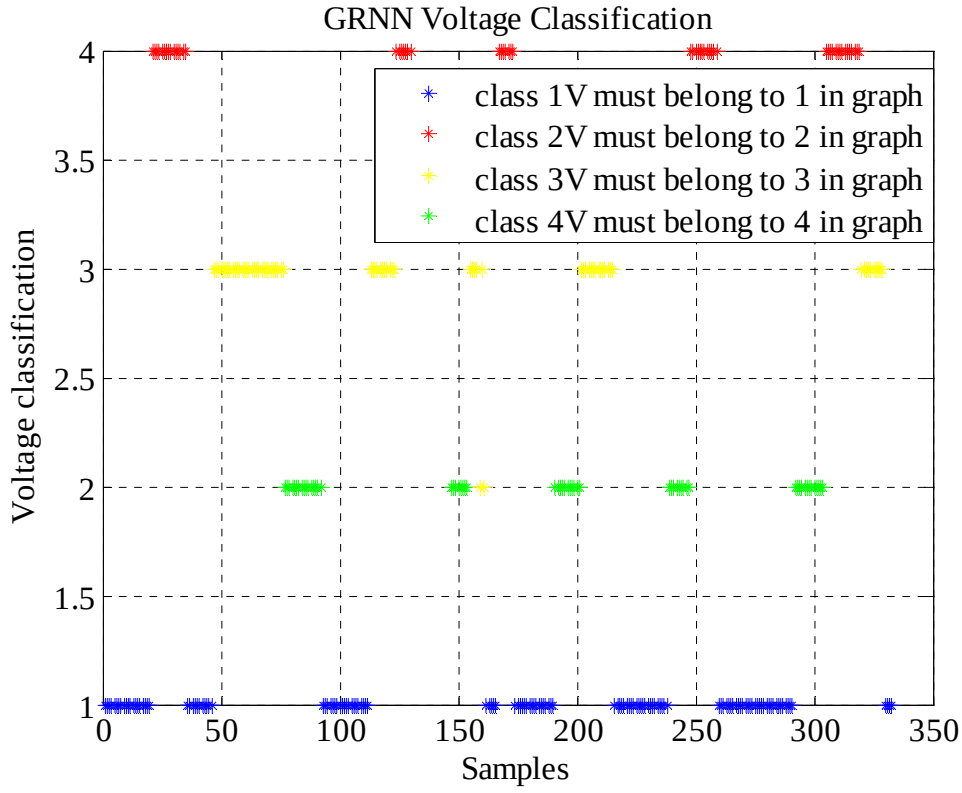


Figure 4.18. Vmpp classification using GRNN.

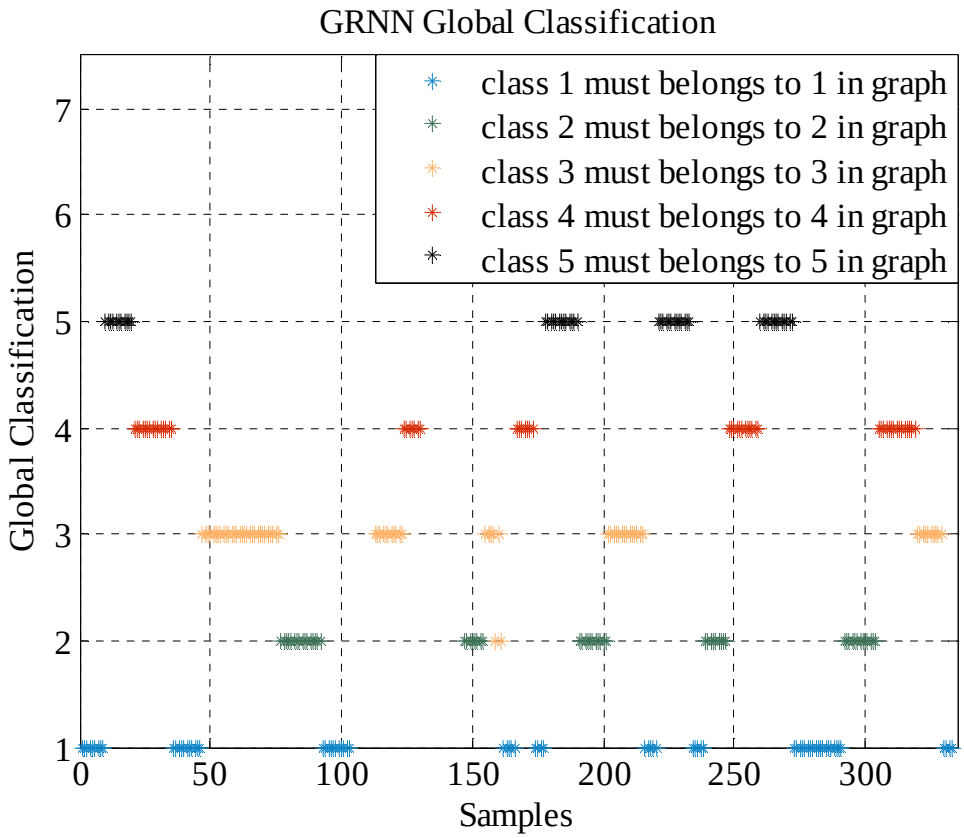


Figure 4.19. Global diagnosis of the system using GRNN.

4.4.5. Training phase and response time

The table below (Table 4.4) resumes the behavior of the five algorithms for the two ANNs representing the five function using in training phase of PV systems diagnosis relative to the present study with their different parameters such as: function, activation function, training spread, training goal and number of neurons in each hidden layer for the two electrical parameters.

Table 4.4. Parameters adjustable of each neural network.

	ANN	**RBF1**	**RBF2**	**PNN**	**GRNN**
Impp	**-Activation function:** tangent sigmoid. **-Function:** newff -**Algorithm:** Levenberg Marquart **-two hidden layers:** 8 x 8 neurons respectively -**training goal=** 10^{-4} **-training epochs=** 15	**-Activation function**: exponential **-Function:** newrbe **-training spread**= 29.7	-**Activation function:** exponential **-Function:** newrb -**training goal=** 10^{-3} -**training spread=** 35 -**One hidden layer** with 1600 neurons. -**MSE:** 0.02816	**-Activation function:** exponential -**Function:** newpnn **-training spread=** 1.05	**-Activation function:** exponential **-Function:** newgrnn **-training spread=** 1.05

Table 4.4. Parameters adjustable of each neural network (the following).

	ANN	**RBF1**	**RBF2**	**PNN**	**GRNN**
Vmpp	- **Activation function:** tangent sigmoid. **- Algorithm**: Levenberg Marquart **-Function:** newff -**One hidden layer:** 45 neurons. **-training goal=** 10^{-4} **-training epochs**= 15	**-Activation function:** exponential **-Function:** newrbe -**training spread=** 49.14	**-Activation function:** exponential **-Function:** newrb -**training goal=** 10^{-2} **-training spread=** 4.9 -**One hidden layer:** 50. neurons. **-MSE:** 0.01325	**-Activation function:** exponential **-Function:** newpnn -**training spread=** 7.38	**-Activation function:** exponential -**Function:** newgrnn -**training spread=** 4.9

In order to keep a high quality of diagnosis in PV systems, a very important criterion should be taken in consideration, which plays a significant role in the time of respond, it is time factor. This factor changes from one algorithm to another. If the response time is short with fewer classification errors, it means that the model is perfectly efficient. Table 4.5 recapitulates the variation of running time corresponding to both electrical parameters and global diagnosis of PV systems with the five different algorithms.

Table 4.5. The variation of running time for the five algorithms.

	BPNN	RBF1	RBF2	PNN	GRNN
Time of (Impp)	13s	12s	26min30s	09s	12s
Time of (Vmpp)	16s	28s	01min36s	15s	25s
Time of (Global diagnosis)	34s	50s	28min20s	30s	45s

4.4.6. Analysis the efficiency of the five different algorithms

Table.6 demonstrates the new nomination of the five different algorithms using IFD in the aim to use it in the next figures of four various criterions.

Table 4.6. The nomination of the five algorithms using IFD.

IFD1	IFD2	IFD3	IFD4	IFD5
BPNN	RBF1	RBF2	PNN	GRNN

4.4.6.1. Accuracy

The Table.7 represents current Accuracy, where the five algorithms display excellent results with 100% for both classes meaning that all samples have been classified in their right classes, the results are proved in Figures 4.5, 4.8, 4.11, 4.12 and 4.17. Table 4.8 represents the voltage accuracy. In this case, the PNN is considered as the best classifier with 100% for all classes verified in Figure 3.16, the BPNN, RBF1 and GRNN display two classes voltage under 100% while the RBF2 three classes under 100%.

Table 4.7. Current Accuracy (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1I	100	100	100	100	100
Class 2I	100	100	100	100	100

Table 4.8. Voltage Accuracy (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1V	100	99.70	100	99.70	100
Class 2V	99.40	98.80	99.399	100	100
Class 3V	100	99.40	99.399	100	100
Class 4V	100	100	100	100	100

Table 4.9 shows the global accuracy according to the five classes of this study, where at first the PNN confirm its best accuracy with 100% for all classes, the BPNN represents a better accuracy (99.69%) than RBF2 and GRNN (99.39%) with two confusion classes while the RBF1 effects the rather good classifier with three classes under 100% accuracy.

Table 4.9. Global Accuracy (%)

	BPNN	RBF1	RBF2	PNN	GRNN
Class 1	99.70	100	99.70	100	100
Class 2	100	99.40	98.80	100	99.40
Class 3	100	100	99.40	100	99.40
Class 4	100	100	100	100	100
Class 5	100	100	100	100	100

Figure 4.20 shows the Confusion of global accuracy according to the five classes of this study, where at first the PNN confirm its best accuracy with 0% confusion for all classes, the BPNN represents a better accuracy between 0-0.15 % than GRNN (0-0.301 %) with two confusion classes. In other hand, the Confusion of accuracy in RBF1 is between (0-0.301 %) which represents disorientation in three classes while the RBF1 effects the rather good classifier with confusion accuracy varying between (0-0.451 %) four classes.

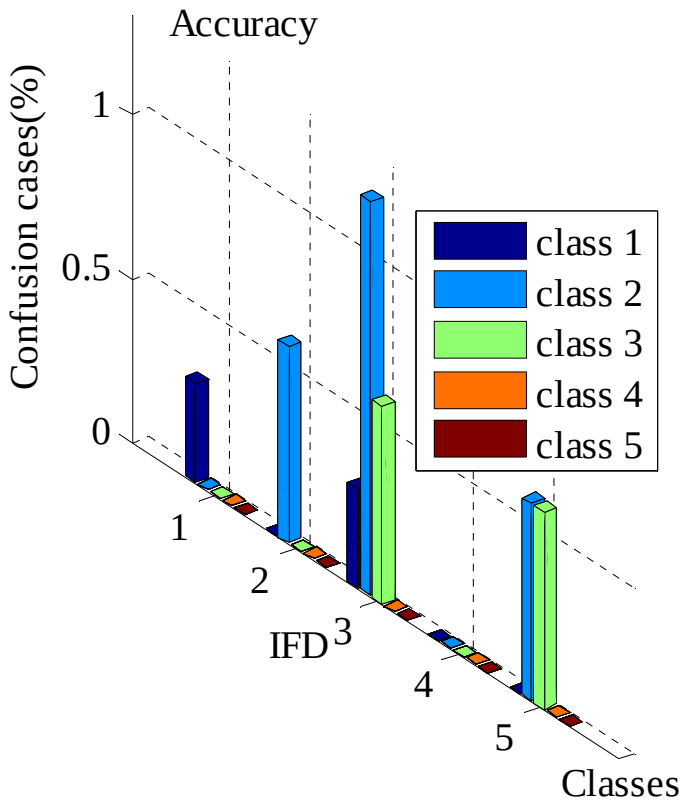


Figure 4.20. Confusion of the global accuracy in the five different algorithms.

4.4.6.2. Sensivity

Table 4.10 represents the same notes cited in the first paragraph of 4.4.6.1. section. While Table 4.11 and Table 4.12 illustrate the voltage and global sensivity respectively. Whereas the PNN presents 100% for all classes, in RBF1, BPNN and GRNN one class under 100% is presented in one PV module short circuited and two PV modules short circuited respectively. The voltage sensivity for RBF2 contains two classes under 100% is presented in one PV module short circuit and two PV modules short circuit, where Figure 4.21 perfectly illustrates the confusion of global sensivity with five algorithms under five different classes.

Table 4.10. Current sensivity (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1I	100	100	100	100	100
Class 2I	100	100	100	100	100

Table 4.11. Voltage Sensivity (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1V	97.70	98.84	100	100	100
Class 2V	100	96.92	100	98.43	100
Class 3V	100	100	96.92	100	100
Class 4V	100	100	100	100	100

Table 4.12. Global Sensivity (%)

	BPNN	RBF1	RBF2	PNN	GRNN
Class 1	100	97.70	98.84	100	100
Class 2	98.43	100	96.92	100	96.92
Class 3	100	100	100	100	100
Class 4	100	100	100	100	100
Class 5	100	100	100	100	100

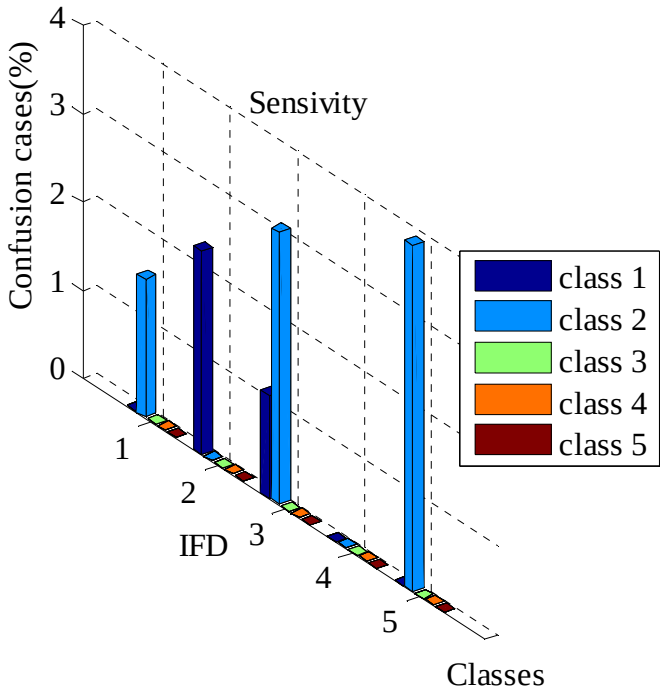


Figure 4.21. Confusion of the global sensivity in the five different algorithms.

4.4.6.3. Precision

Table 4.13 represents the same notes cited in the first paragraph of 4.4.6.1. section.

Table 4.13. Current Precision (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1I	100	100	100	100	100
Class 2I	100	100	100	100	100

Table 4.14. Voltage Precision (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1V	100	100	100	98.82	100
Class 2V	96.82	98.41	100	100	100
Class 3V	100	97.22	97.22	100	100
Class 4V	100	100	100	100	100

Table 4.15. Global Precision (%)

	BPNN	RBF1	RBF2	PNN	GRNN
Class 1	98.82	100	100	100	100
Class 2	100	96.82	98.41	100	100
Class 3	100	100	97.22	100	97.22
Class 4	100	100	100	100	100
Class 5	100	100	100	100	100

The excellent results for precision are in both algorithms (PNN and RBF1) with 100% for all classes, the BPNN and GRNN manifest one class under 100% specifically in one PV module short circuited, while the RBF2 reveals its poor precision with three classes under 100% shown in Table 4.14 and Table 4.15 respectively and better prove in Figure 4.22.

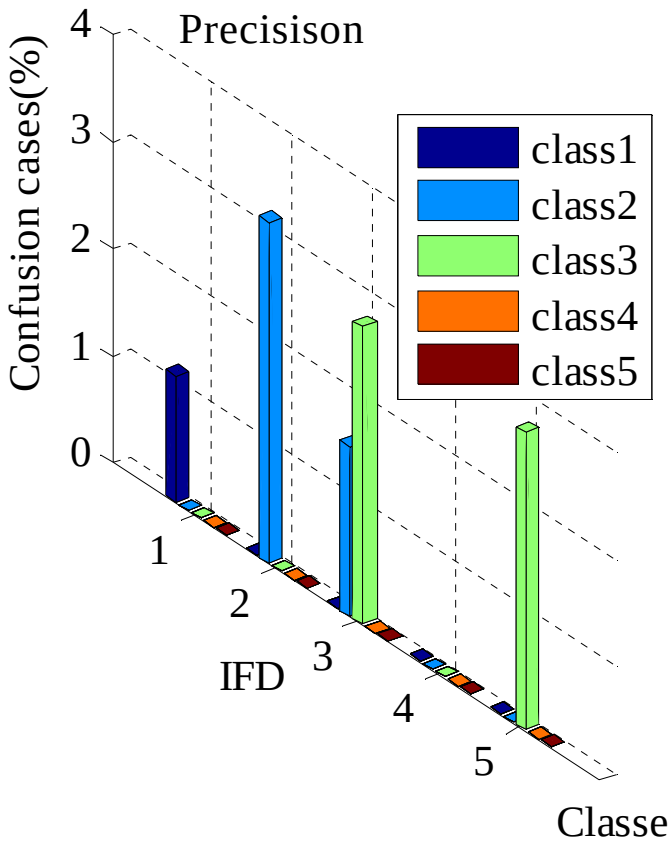


Figure 4.22. Confusion of the global precision in the five different algorithms.

4.4.6.4. Specificity

Table.16 represents the same notes cited in the first paragraph of Section 4.4.6.1. The best result for the specificity is in the PNN algorithm displaying 100% for all classes, the rather good result is represented in the BPNN and GRNN algorithms respectively with one class under 100% demonstrated in Table.16 and Table.17. The RBF1 and RBF2 reflect their very insufficient specificity displaying two and three classes respectively under 100%. Where the result is illustrated in Figure 4.23.

Table 4.16. Current Specificity (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1I	100	100	100	100	100
Class 2I	100	100	100	100	100

Table 4.17. Voltage Specificity (%)

	RBF1	RBF2	GRNN	BPNN	PNN
Class 1V	100	100	100	100	100
Class 2V	99.26	99.69	100	100	100
Class 3V	100	99.24	97.29	97.29	100
Class 4V	100	100	100	100	100

Table 4.18. Global Specificiy (%)

	BPNN	RBF1	RBF2	PNN	GRNN
Class 1	100	100	100	100	100
Class 2	100	99.26	99.629	100	100
Class 3	97.29	99.69	99.24	100	97.29
Class 4	100	100	100	100	100
Class 5	100	100	100	100	100

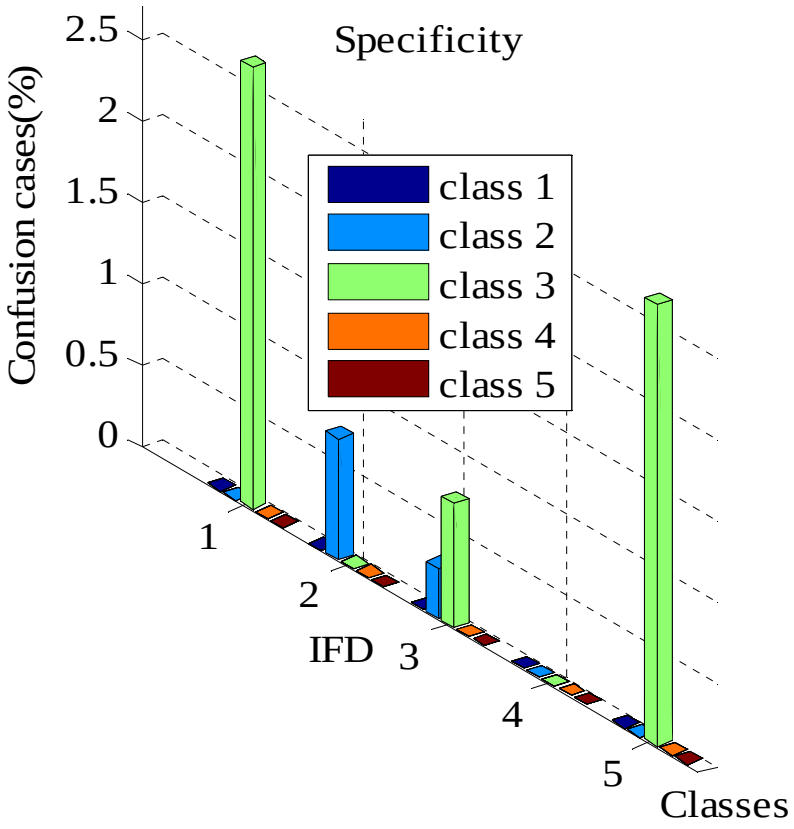


Figure 4.23. Confusion of the global specificity in the five different algorithms.

Figure 4.24 illustrates confusion of the four criterions (accuracy, sensivity, precision and specificity) in the five different algorithms noting that the highest confusions focus in the cases of one PV module short circuited and two PV modules short circuited, some confusion in healthy system and faulty string, while four PV modules short circuited do not represent any confusion in all algorithms.

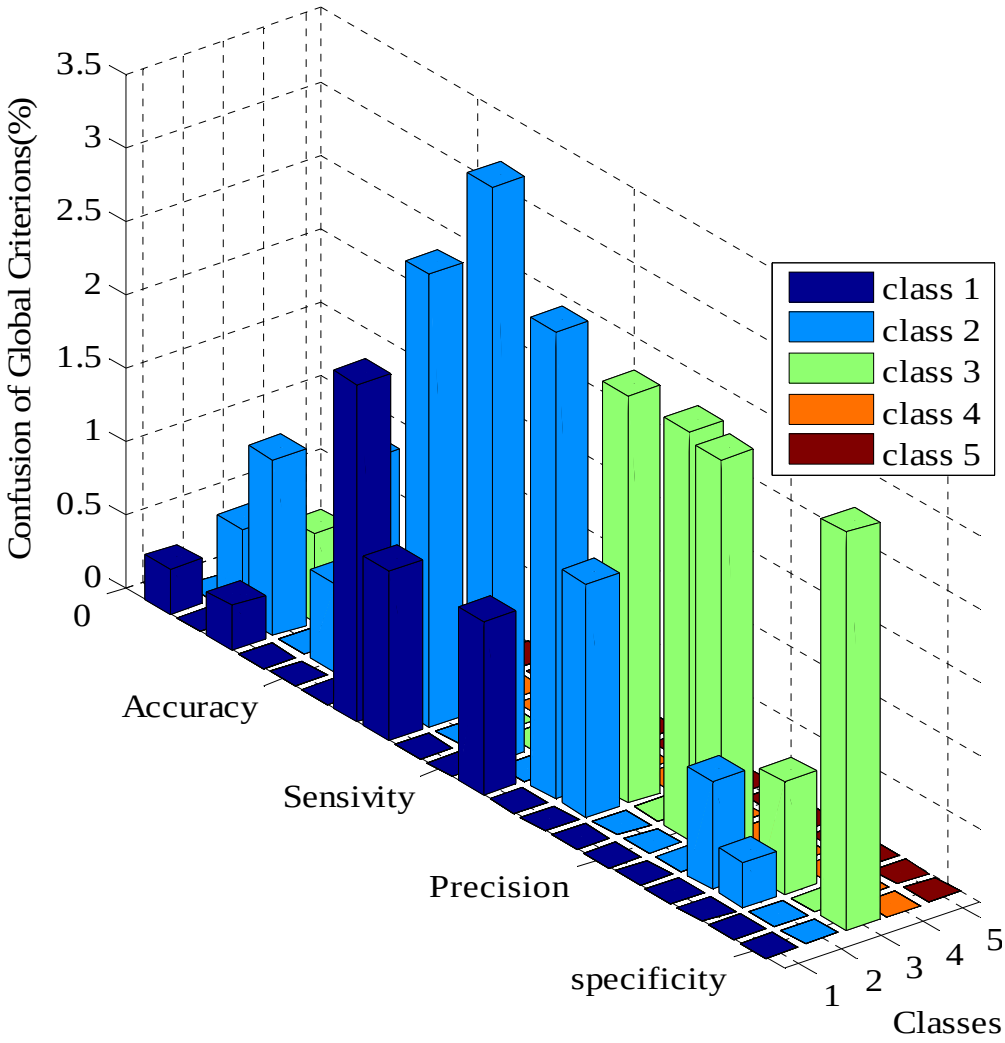


Figure 4.24. Confusion of the global criterions in the five different algorithms.

Table 4.7 specifies the localization of confusion in all algorithms, where the confusion is between healthy system and one PV module short circuit and between one PV module short circuit and two PV modules short circuit. In these cases, the confusions are due to the variation of the temperature.

Table 4.19. Localization of confusions in all algorithms.

	CONFUSIONS
BPNN	One sample of healthy system in one PV module short circuit.
RBF1	Two samples of one PV module short circuit in healthy system.
RBF2	- One sample of One PV module short circuit in healthy system. - Two samples of two PV modules in one PV module short circuit.
PNN	No confusion for all samples in all classes.
GRNN	Two samples of two PV modules short circuit in one PV module short circuit.

4.5. Conclusion

This chapter has brought up comparative study of fault detection and diagnosis of PV generator with five different neural networks. The algorithms can locate and classify the frequent fault encountered in PV system such as: disconnected branch in PV generator and various cases of short circuit fault. The efficiency and reliability of these algorithms has been tested under four principal criterions called key statistical concept that are: accuracy, sensivity, specificity and precision in order to find the best classifier algorithm. Additionally, this approach has taken another important factor in consideration playing a significant role to keep a high quality of diagnosis, which is time of respond.

The results mark a good illustration from point of view of classification with tiny confusions in some algorithms between healthy, one PV module short circuited and between one PV module short circuited and two PV modules short circuited. In accordance with efficiency and response time of these algorithms: the lowest efficient algorithm is RBF2 as its performance varies between 96.82 and 100% with three confusion classes and its response time is around 28 minutes. While, the satisfactory algorithms are RBF1 and GRNN with two confusion classes in short response time followed by BPNN algorithm which represents a very well results of performance and at the end, the PNN algorithm reveals its high quality of performance displaying 100% in all key statistical concept with a shortest response time related to the others algorithms.

General Conclusion

Over the last few years, modern civilisation thinks and discusses profoundly about energy transition from fossil fuel to renewable energies especially photovoltaic energy. The aim of this conversion focuses not only in the impact of climate change but also in the impact of public and individual health due to increasing morbidity and mortality.

However, the production of photovoltaic energy growing up in staggering rate for its cleanliness, safety, quiet, reduction in energy bills and low maintenance requirement. With a view to achieve a maximum harvest of energy supplied by PV systems, diagnosis presents an indispensable and crucial tool to maximize power production, reliability, efficiency, safety and quality in global PV systems.

In the present thesis, a new intelligent diagnosis solution is presented. This solution uses one type of Artificial Intelligence which is the Artificial Neural Network. Thereunto, the choice of ANN has been done for its simplicity, flexibility, established usage as well as its rapidity. The developed method for PV grid connected systems' fault diagnosis requires two ANNs that classify separately the current and the voltage of maximal power point respectively. The outputs of both ANNs are the input of a combinational algorithm where its output provides the global diagnosis of PV generator. The global diagnosis of the studied PV system was able to prove its quality by its excellent accuracy and efficiency from the point of view of fault localization in accordance to an average overall accuracy of 98.6%.

The last part in this thesis examines the impact of the Artificial Neural Network choice on the performance of an IFD diagnosis algorithm, which is designed to detect and isolate faults in grid connected PV installation. Four most pertinent criterions are considered in this analysis: the accuracy, the specificity, the sensitivity and the rapidity. The goal is to choose the best ANN to ensure the IFD diagnosis efficiency and consequently PV system's safety, durability and reliability. For this propose, five different

ANNs have been used and tested: Back Propagation Neural Network, Probabilistic Neural Network, Generalized Regression Neural Network and two Radial Basis Function Neural Network. The performances of these ANNs have been analyzed and compared. These five Neural Networks have the same inputs as the first part of the work: cell temperature, solar irradiation, voltage and current of the maximal power point of the I-V characteristics. The same faults treated in the last ANN (BPNN) as well as the same working conditions.

For their comparison, the five ANNs based IFD algorithms has been tested using 336 different functioning cases enclosing healthy functioning cases and different faulty functioning cases. These investigations demonstrate that the RBF2 based algorithm presents the lowest efficiency with a response time equal to 28 min. noting that its performance varies from 96.82% to 100% on all criterions with three confusion cases for faults classification. While the obtained results from GRNN and RBF1 have presented good results on all criterions with a short response time and a good classification with two confusion classifications cases. Concerning the BPNN based IFD algorithm, the tests reveal very good results on all criterions varying from 97.27% to 100% with a very good classification’s score (one confusion sample). Finally, the PNN based IFD algorithm displaying 100% of success score on all key statistical concepts compared to other ANNs with no confusion cases. This places it at the top of the list and qualifies it as the best intelligent diagnosis algorithm for the studied grid connected PV installation.

The efficiency of the developed method is experimentally evaluated by using real measured data, collected from real Grid Connected PV system is part of the Algerian-Spanish cooperation placed on the roof and is composed by 90 monocrystalline PV modules provides output power equal to 9.54 kWp located at the Centre Des Energies Renouvelables in Algiers (Algeria).

In this study, the focus was on the fault detection and diagnosis of PV generator in small grid connected PV system. For that, five algorithms using ANNs were studied

and analyzed. In order to guarantee perfect and more accurate methods of fault detection and diagnosis of PV systems, more advanced methods are introduced by the researchers for numerous purposes. Therefore, a significant future work is investigating other of AI and signal processing techniques for many failures cases in PV systems installations at a high scale in real time. In addition, improved AI fault diagnosis techniques can be applied in a more in-depth way from generator to PV cell by integrating artificial intelligence hybridization for certain fault cases requiring a high precision. In other words, the used of technics citing above in fault detection and diagnosis can be not only in PV stations but also in smart grid, electric vehicle or in electrical machines employing PV systems.

REFERENCES

[1] Birol F. Solar PV, on track. International Energy Agency; 2020.

[2] https:// www.sktm.dz/

[3] Mellit. A, Tina.G.M., Kalogirou .S.A. Fault detection and diagnosis methods for photovoltaic systems: A review.Ren.Sustain Energy Rev 2018; 91:1-18.

[4] Y. Hu, B. Gao, X. Song, G. Tian, K. Li, X. He. Photovoltaic fault detection using a parameter based model. Sol. Energy 2013; 96: 96-102.

[5] Harrou F, Fillatre L, Nikiforov I. Anomaly detection/detectability for a linear model with a bounded nuisance parameter. Ann. Rev. Contr 2014; 38 (1):32–4.

[6] M Dhimish, V Holmes. Fault detection algorithm for grid-connected photovoltaic plants. Sol. Energy 2016; 137: 236-245.

[7] Chouder A, Silvestre S. Automatic supervision and fault detection of PV systems based on power losses analysis. Energy Convers. Manag 2010; 51: 1929-37.

[8] Gokmen N, Karatepe E, Celik B, Silvestre. Simple diagnostic approach for determining of faulted PV modules in string based PV arrays. Sol. Energy 2012; 86: 3364–77.

[9] Chouder A, Silvestre S. Analysis model of mismatch power losses in PV systems. Sol. Energy 2009; 131 (2):024504-4.

[10] Silvestre S, Chouder A, Karatepe E. Automatic fault detection in grid connected PV systems. Sol. Energy 2013; 94: 119–27.

[11] Ammiche M, Kouadri A, Halabi L-M, Guichi A, Mekhilef S. Fault detection in a grid-connected photovoltaic system using adaptive thresholding method. Sol. Energy 2018; 174: 762–69.

[12] Drews A, de Keizer A.C, Beyer H.G, Lorenz E, Betcke J, van Sark W.G.J.H.M, Heydenreich W, Wiemken E, Stettler S, Toggweiler P, Bofinger S, Schneider M, Heilscher G, Heinemann D. Monitoring and remote failure detection of grid connected PV systems based on satellite observations. Sol. Energy 2007; 81: 548–64.

[13] Chine W, Mellit A, Pavan A M, Kalogirou S A. Fault detection method for grid-connected photovoltaic plants. Renew. Energy 2014 ; 66: 99–10.

[14] Platon R, Martel J, Woodruff N, Chau T.Y. Online fault detection in PV systems. IEEE Trans. Sustain. Energy 2015; 6 (4):1200–07.

[15] Bastidas-Rodriguez J.D, Franco E, Petrone G, Ramos-Paja C.A, Spagnuolo G. Quantification of photovoltaic module degradation using model based indicators. Mathematics and Computers in Simulation 2017, 131:101-13.

[16] Dhoke A, Sharma R, Kumar Saha T. An approach for fault detection and location in solar PV systems. Sol. Energy 2019; 194: 197-08.

[17] Garoudja E, Harrou F, Sun Y, Kara K, Chouder A, Silvestre S. Statistical fault detection in photovoltaic systems. Sol. Energy 2017; 150: 485-99.

[18] Chen Z, Chen Y, Wu L, Cheng S, Lin P. Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. Energy Convers. Manage 2019; 198: 111793.

[19] Mansouri M, Al-khazraji A, Hajji M, Harkat M-F, Nounou H, Nounou M. Wavelet optimized EWMA for fault detection and application to photovoltaic systems. Sol. Energy 2018; 167: 125-36.

[20] H Wang , J Zhao, Q Sun, H Zhu. Probability modeling for PV array output interval and its application in fault diagnosis. J Energy 2019. 189: 116248.

[21] F Harrou, Y Sun, B Taghezouit, A Saidi, M-Ek Hamlati. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. Ren. Energy 2018. 116: 22-37.

[22] Takashima T, Yamaguchi J, Otani K, Oozeki T, Kato K, Ishida M. Experimental studies of fault location in PV module strings. Sol Energy Mater & Sol Cel 2009; 1079–82.

[23] Solórzano J, Egido M.A. Automatic fault diagnosis in PV systems with distributed MPPT. Energy Convers. Manage 2013; 76: 925–34.

[24] Yongjie Liu, Kun Ding, Jingwei Zhang, Yuanliang Li, Zenan Yang, Wenming Zheng, Xiang Chen. Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with I-V curves. Energy Convers. Manage 2021; 245: 114603.

[25] Jingna Pan, Weiguo He, Yucheng Shi, Ruyin Hou, Honglu Zhu. Uncertainty analysis based on non-parametric statistical modelling method for photovoltaic array output and its application in fault diagnosis. Sol. Energy 2021; 225 : 831–41.

[26] Eskandari A, Milimonfared J, Aghaei M. Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics. Sol. Energy 2020; 211 : 354–65.

[27] Spataru S, Sera D, Kerekes T, Teodorescu R. Diagnostic method for photovoltaic systems based on light I–V measurements. Sol. Energy 2015; 119 : 29–44.

[28] Hachana O, Giuseppe Marco Tina, Hemsas K. PV array Fault Diagnostic Technique for BIPV Systems, Energy & Buildings 2016; 126: 263-74.

[29] SolarEdge, 2013. Performance of PV Topologies under Shaded Conditions (white paper). SolarEdge.

[30] Swingler A, 2010. Photovoltaic String Inverters and Shade-Tolerant Maximum Power Point Tracking: Toward Optimal Harvest Efficiency and Maximum ROI (white paper). Schneider Electric, Burnaby, Canada.

[31] Mingyao M, Zhixiang Z, Zhen X, Ping Y, Xing Z, Fei L. Fault diagnosis of cracks in crystalline silicon photovoltaic modules through I-V curve. Microelectronics Reliability 2020; 114: 113848.

[32] Chaibi Y, Malvoni M, Chouder A, Boussetta M, Salhi M. Simple and efficient approach to detect and diagnose electrical faults and partial shading in photovoltaic systems. Energy Convers. Manage 2019; 196: 330-343.

[33] Mekki H, Mellit A, Salhi H. Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules. Simulation Modelling Practice and Theory 2016; 67:1-13.

[34] Shrikhande S, Varde P, Datta D. Prognostics and health management: methodologies & soft computing techniques. In: Current Trends in Reliability, Availability, Maintainability and afety. Springer 2016; 213–27.

[35] Hare J, Shi X, Gupta S, Bazzi A. Fault diagnostics in smart micro-grids: a survey. Renew. Sustain. Energy Rev 2016; 60:1114–24.

[36] Suganthi L, Iniyan S, Samuel A. Applications of fuzzy logic in renewable energy systems – a review. Renew. Sustain. Energy Rev 2015; 48: 585–07.

[37] Silvestre S, daSilva M, Chouder A, Guasch D, Karatepe E. New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators. Energy Convers. Manage 2014; 86:2 41–49.

[38] Tadj M, Benmouiza K, Cheknane A, Silvestre S. Improving the performance of PV systems by faults detection using GISTEL approach. Energy Convers. Manage 2014; 80:298–04.

[39] Zhao Y, dePalma J, Mosesian J, Lyons R, Lehman B. Line–line fault analysis and protection challenges in solar photovoltaic arrays. IEEE Trans. Ind. Electron 2013; 60 (9):3784–95.

[40] Li Z, Wang Y, Zhou D, Wu C. An intelligent method for fault diagnosis in photovoltaic array, ICSC Part II CCIS 2012;327 :10-6.

[41] Siva Ramakrishna Madeti, S.N. Singh. Modeling of PV system based on experimental data for fault detection using kNN method. Sol. Energy 2018; 173: 139–151.

[42] M Borunda, O.A.Jaramillo, A Reyes, Pablo H.Ibargüengoytia. Bayesian networks in renewable energy systems: A bibliographical survey. Renew. Sustain. Energy Rev 2016; 62, 32-45.

[43] Bonsignore L, Davarifar M, Rabhi A, Tina G.M, Elhajjaji A. Neuro-Fuzzy fault detection method for photovoltaic systems. Energy Procedia 2014;62 : 431-41.

[44] Dhimish M, Holmes V, Mehrdadi B, Dales M. Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection. Renew. Energy 2018; 117: 257–274.

[45] Dhimish, M., Holmes, V., Mehrdadi, B., Dales, M.et Mather, P. Photovoltaic fault detection algorithm based on theoretical curves modelling and fuzzy classification system. J. Energy 2017;140: 276-290.

[46] Chine W, Mellit A, Lughi V, Malek A, Sulligoi G, Massi Pavan A. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renew Energy 2016; 90: 501-512.

[47] Hussain M, Dhimish M, Titarenko S, Mather P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. Renew. Energy 2020; 155: 1272–92.

[48] Kara Mostefa Khelil C, Amrouche B, Benyoucef AS, Kara K, Chouder A. New Intelligent Fault Diagnosis (IFD) Approach for grid-connected photovoltaic systems. J Energy 2020;211. 118591.

[49] Zhenghai L, Dazheng W, Liangliang T,Jinli R, Zhuming L. A Heuristic Diagnostic Method for a PV System: Triple-Layered Particle Swarm Optimization–Back-Propagation Neural Network, Energies 2017; 10:226.

[50] Garoudja E, Chouder A, Kara K, Silvestre S. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. En Con Man 2017; 151: 496–13.

[51] Zhu H, Lub L, Yao J, Daia S, Hu Y. Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. Sol. Energy 2018; 176: 395–05.

[52] R Benkercha, S Moulahoum. Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system. Sol. Energy 2018; 173: 610–34.

[53] Gompel.J.V, Spina.D, Develder.C. Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks. J.apenergy 2021; 305:117874.

[54] Xiaoyang Lu, Peijie Lin, Shuying Cheng, Gengfa Fang, Xiangjian He, Zhicong Chen, Lijun Wu. Fault diagnosis model for photovoltaic array using a dual-channels convolutional neural network with a feature selection structure. En Con Man 2021; 248: 114777.

[55] Hajji M, Harkat M-F, Kouadri A, Abodayeh K, Mansouri M, Nounou H, Nounou M. Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems. European J Cont 2021. 59: 313-21.

[56] Li B , Delpha C , Diallo D, Migan-Dubois A. Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review. Renew. Sustain. Energy Rev 2021; 138, 110512.

[57] Kara Mostefa Khelil C, Amrouche B, Kara K, Chouder A. The impact of the ANN's choice on PV systems diagnosis quality. En Con Man 2021; 240: 114278.

[58] Burnham L. Renewable Energy: Sources for Fuels and Electricity. Island Press 1993: 1160.

[59] Kashife I, Zainal S. A review of maximum power point tracking techniques of PV system for uniform insolation and partial shading condition. Renew. Sustain. Energy Rev 2013; 19, 475-88.

[60] Kichou, S., Abaslioglu, E., Silvestre, S., Nofuentes, G., Torres-Ram.rez, M.et Chouder, A. Study of degradation and evaluation of model parameters of micromorph silicon photovoltaic modules under outdoor long term exposure in Jaén, Spain. En Con Man, V. 120, (2016), 109-119.

[61] Mathworks https://fr.mathworks.com/products/simscape.html. Modélisation et simulation de systèmes physiques multi-domaines. Consulté le jeudi 07 juin 2018.

[62] Amrouche B, Guessoum A, Belhamel M. A simple behavioural model for solar module electric characteristics based on the first order system step response for MPPT study and comparison. Applied Energy 2012; 91(1):395-04.

[63] Amrouche B. Improvement and experimental validation of a simple behavioural model for photovoltaic modules. Solar Energy Materials and Solar Cells 2014; (128):204-14.

[64] Ahmadipour M, Hizam H, Lutfi Othman M, Moh Radziab M, Srikanta Murthy A. Islanding detection technique using Slantlet Transform and Ridgelet Probabilistic Neural Network in grid-connected photovoltaic system., j apenergy 2018., 231: 645-59.

[65] Amrouche B, Pivert X. Artificial neural network based daily local forecasting for global solar radiation. J Appl Energy 2014;130:333–41.

[66] Karatepe E, Hiyama T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In: 2011 16th international conference on intelligent system application to power systems (ISAP). IEEE; 2011. p. 1–6.

[67] Chun F, Nanjing Z, Gaofang Y, Tingting G, Ruifang Y, Xiaowei C, Min C, Jingbo D. Artificial neural networks combined multi-wavelength transmission spectrum feature extraction for sensitive identification of waterborne bacteria. J Spectr Part A: Mol and Biom Spect 2021: (251) 119423.

[68] Liang L, Guo W , Z Yu, Z WeiZhe , Li L, Xing X. Radial Basis Function Neural Network for prediction of medium frequency sound absorption coefficient of composite structure open-cell aluminum foam. Applied Acoustics 2020: (170)107505.

[69] Matera Fabio. Radial basis function neural network. Subst Use Misuse 1998;33 (2):317–34.

[70] Hong H., Zhang Z, Guo A, Shen L, Sun H, Liang Y, Wu F, Lin H., Radial basis function artificial neural network (RBF ANN) as well as the hybrid method of RBF ANN and grey relational analysis able to well predict trihalomethanes levels in tap water., J Hydrology (2020)125574.

[71] Chine W. Contribution au diagnostic des défauts dans les systèmes photovoltaïques. Doctoral thesis(2017). Jijel university, Algeria.

[72] https://fr.mathworks.com/help/deeplearning/ref/newrb.html.

[72] Specht DF. Probabilistic neural networks for classification, mapping, or associative memory. IEEE international conference on neural networks. 1988. p. 525–32.

[73] H.Fan, J.Pei, Y.Zhao. An optimized probabilistic neural network with unit hyperspherical crown mapping and adaptive kernel coverage. Neurocomputing 2020., 373: 24-34.

[74] Donald F. Specht., A General Regression Neural Network., IEEE TRANSACTIONS ON NEURAL NETWORKS. VOL.2. NO. 6. NOVEMBER 1991.

[75] Q. Yuan, H. Xu, T. Li, H. Shen, L. Zhang. Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental U.S., J Hydrology 2020., (580)124351.

[76] Ting L, Chudong T, Haizhen Y, Xuhua S, Lijia L. Nonlinear process monitoring based on decentralized generalized regression neural networks. J exp syst with appl 2020., (150)113273.

[77] H.L.Bank, M.K.Schmehl. Parameters for evaluation of viability assays: Accuracy, precision, specificity, sensitivity, and standardization., Cryobiology 1989. (26)203-11.

[78] David Martin Ward Powers. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation., J Machine Learning Technologies 2011. (2)37-67.

[79] Garoudja E. FAULTS DETECTION AND DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS USING ARTIFICIAL INTELLIGENCE TOOLS. Doctoral thesis (2018). Blida 1 University, Algeria.

APPENDIX ‘A’

First Publication

Kara Mostefa Khelil, C ., Amrouche, B., Benyoucef ,AS., Kara, K., Chouder, A. (2020).New Intelligent Fault Diagnosis (IFD) approach for grid-connected photovoltaic systems. J Energy, 211, 118591.

Source Normalized Impact Per Paper (SNIP): 2.014

SCImago Journal Rank (SJR): 1.961

Impact Factor: 7.147 (2020 Journal Citation Reports)

5 Year Impact Factor: 6.845 (2020 Journal Citation Reports)

Cite Score: 11.5

Energy 211 (2020) 118591

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

New Intelligent Fault Diagnosis (IFD) approach for grid-connected photovoltaic systems

Chérifa Kara Mostefa Khelil [a, b, *], Badia Amrouche [a, b], Abou soufiane Benyoucef [a, b], Kamel Kara [a, b], Aissa Chouder [a, b]

[a] SET Laboratory, Electronics Department, Blida 1 University, BP 270, Blida, Algeria
[b] Electrical Engineering Laboratory (LGE), University Mohamed Boudiaf of M'sila, BP 166, 28000, Algeria

ARTICLE INFO

Article history:
Received 16 March 2020
Received in revised form
30 June 2020
Accepted 10 August 2020
Available online 19 August 2020

Keywords:
Photovoltaic energy
PV grid Connected power plant
Faults
Diagnosis
Intelligent fault diagnostic
Artificial neural networks

ABSTRACT

The present work brings a new intelligent algorithm for PV system's diagnosis and fault detection (IFD). At this stage of the study, this algorithm can detect and identify three recurrent cases between healthy and short circuit faults, as well as string disconnection in PV array using artificial neural networks (ANN). Both, detection and isolation are simple and fast. The developed model requires small training period and is based on only four inputs: the maximum power current and voltage from the output current-voltage (I −V) characteristic, the solar irradiation and the cell temperature. Experimental validation of the proposed IFD has been carried on small grid connected PV generator (PVG). The obtained results demonstrate that this approach can precisely detect and classify the existing faults with high accuracy (98.6%).

1. Introduction

In the last few decades, renewable energies have gained great importance around the world, their development increases exponentially to the detriment of conventional energy sources such as: oil, coal, lignite and natural gas which are the original source of climate change phenomenon, environmental pollution, human health degradation and the growth of the greenhouse effect [1]. Furthermore, renewable energies, in particular photovoltaic energy is safe and clean, with no carbon di-oxide emission, widely available and environment-friendly. The growth of photovoltaic energy production will have a very important role in the electricity production in the future. It has been proven that solar irradiance is the biggest source nowadays, it has achieved a cumulative capacity of 137.5 Giga Watt were newly installed in 2019 with a forecast of 750 Giga Watt by end of 2020 [2].

To achieve the best power-generation efficiency, PV systems must work under particular conditions; inshaded area, high irradiation level, low temperature and PV panel optimal orientation. In addition, PV panels must be clean because the accumulation of dirt masks the solar irradiation and reduces the efficiency of the global system. This necessitates regular maintenance in addition to the monitoring. In fact, during their long lifetime, of about 25 years, PV systems can be the subject of numerous faults. That explains the importance of the diagnosis and fault detection for PV systems which is necessary not only to increase system power generation reliability but also for operating costs reduction. Real-time diagnosis of PV system has drawn many researchers attention nowadays. As a result, a great deal of recent research on PV systems have been focusing on that area to help to make possible the fault detection and isolation especially for the PV generator.

On the basis of a literature overview, it was concluded that PV diagnosis techniques can be classified in a number of categories. The first category includes model based diagnosis techniques [3] in which the model of the system is used to decide about the occurrence of faults. For this, the simulation values are compared to the system outputs. The inputs of the model are mainly the meteorological working conditions and the electrical parameters of the installed PV modules [4–9,12]. The output of the model is the maximum power point of the PV generator. Additional parameters have been used while dealing with grid connected PV power plants

* Corresponding author. SET Laboratory, Electronics Department, Blida 1 University, BP 270, Blida, Algeria.
E-mail addresses: c.karamostefa@univ-blida.dz, karasolar@yahoo.fr (C. Kara Mostefa Khelil).

https://doi.org/10.1016/j.energy.2020.118591
0360-5442/

such as DC input/output ratio, AC input/output ratio and reference yield measurements [10,11]. Model-based fault detection methods use residuals which indicate changes between the PV generator and its model. This difference is used to estimate the increase of the series resistance and the decrease of the parallel resistance using fill factor method [13], and can be used to focus on intra-string line-line fault [14] or to detect the shading using exponentially weighted moving average estimation (EWMA) [15]. Another method to detect and diagnose faults such: short circuit, open circuit, shading and degradation in PV arrays by residuals using current-voltage curves and ambient conditions based on deep neural network has been also developed [16].

In the second category, are classified the signal processing based fault diagnosis methods. This includes methods based on mathematical or statistical analysis such as time domain reflectometry [17] and Fourier analysis [18], which are used to identify

Table 1
Electrical properties of the Isofoton 106–12 PV module [14,32].

Solar Panel electrical characteristics	Value
Peak power	106 W
Short circuit current (Isc)	6.54 A
Open circuit voltage (Voc)	21.6 V
Voltage at Maximum Power Point (Vmpp)	17.4 V
Current at Maximum Power Point (Impp)	6.10 A
Number of cells connected in Series	36
Number of cells connected in Parallel	2
Cell Short circuit current	3.27 A
Cell Open circuit Voltage	0.6 V

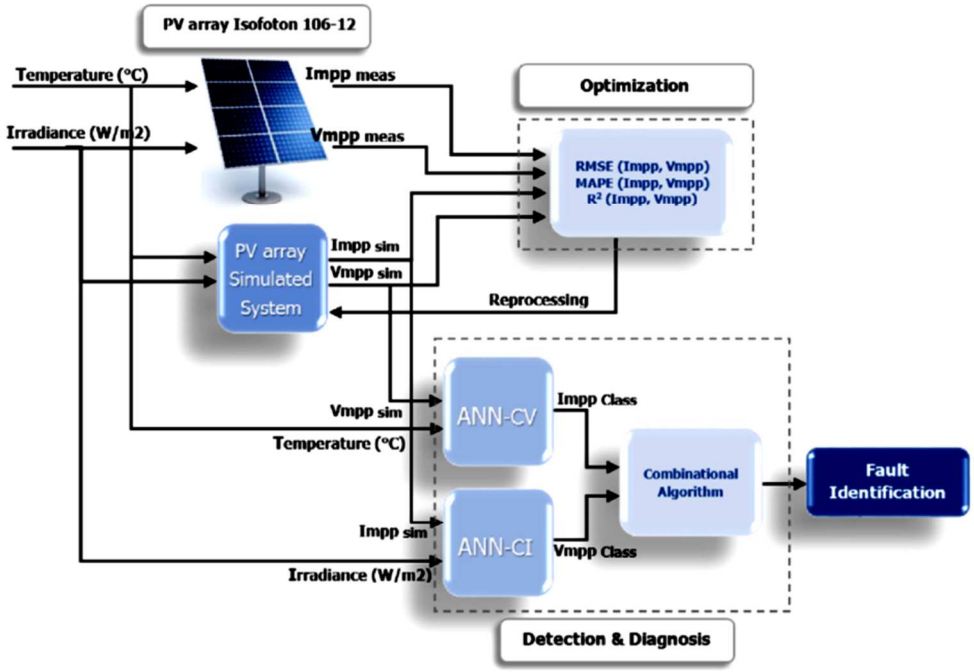


Fig. 1. Schematic description of the IFD methodology.

Fig. 2. Roof grid connected PV plant in Algiers, Algeria.

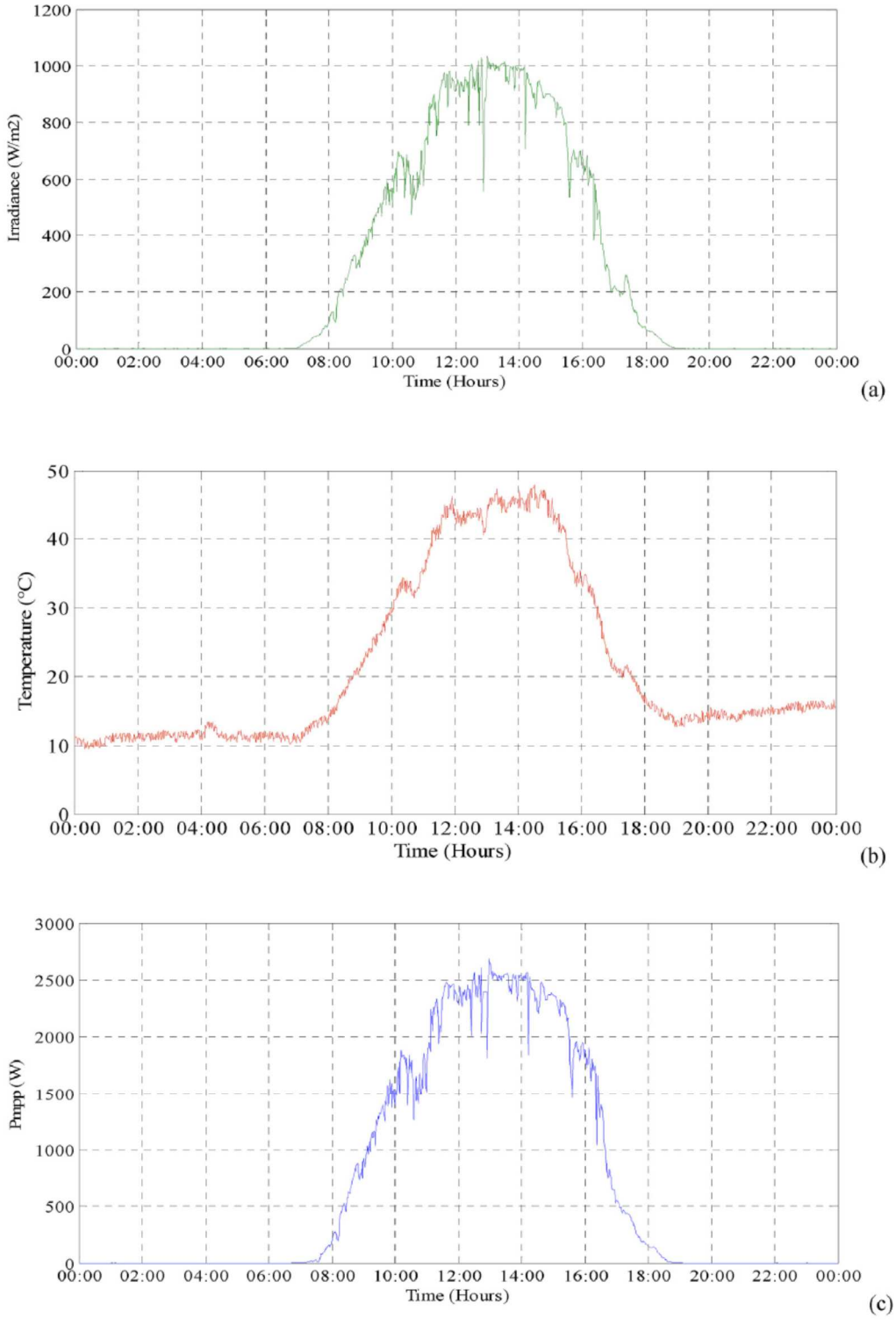


Fig. 3. Daily profile of onsite measured (a) Irradiance, (b) Temperature and (c) Output power.

and locate open circuit faults in grid connected PV plant for example. Whoever, these techniques should be combined to other diagnostic methods as earth capacitance measurement to expand their faults diagnosis capability [19,20]. Measurements are generally the current-voltage (I–V) curve. This curve can provide crucial information such as short-circuit current, open circuit voltage, fill factor indicator, ideal factor, series and shunt resistors values for a PV module, a PVG, a string or branch or PV field [21,22,26,27]. These methods are known to be very efficient to detect and isolate the presence of shading and soiling. Currently, the new generation of PV modules with integrated converters, are fully or partially able to sweep the IV curve of the photovoltaic generator [23–25].

In the third category, are classified computational intelligence-based techniques involving, amongst other concepts, Artificial Neural Networks (ANNs), Fuzzy Logic (FL), Genetic Algorithms (GA) and k-Nearest Neighbors (kNN) [28–38]. In Ref. [39] Bayesian ANNs and polynomial regression method are used to estimate the effect of soiling in large-scale PV systems. While Fuzzy Logic approach has been proposed to detect possible solar panel abnormalities [33,40,41]. In Ref. [42] an interesting comparison is presented between Mamdani Sugeno fuzzy logic and RBF ANN for PV systems to detect faulty module, two faulty modules and partial shading. Other works demonstrate the effectiveness of ANNs which have been successfully used to detect and localise short circuit in PV array [43–45], partial shading, by-pass diode, open circuit, short circuits systems and increased series resistance using the ration information existing between the measured and the simulated data for both current and voltage [28,44,46]. Other PV array Fault Diagnostic Techniques (FDT) have been adopted and validated for Building Integrated PV Systems (BIPV) using Artificial Bee Colony (ABC) algorithm and differential evolution (DE) algorithm method [27]. A Heuristic Diagnostic Method has been proposed for a PV System combining Particle Swarm Optimization (PSO) algorithm and ANNs [47]. In Ref. [48], a probabilistic ANN has been proposed to detect short circuit fault and disconnection fault in the string. The inputs of the ANN are the temperature, the irradiance, the voltage and the current at the MPP. In Ref. [49], a probabilistic ANN has been proposed to diagnose open circuit, short circuit, shading and abnormal aging faults. The inputs of the last ANN are open circuit voltage, short circuit current, current of Maximal Power Point and Voltage of Maximal Power Point. In Ref. [50], a decision tree method has been proposed to detect different types of faults such as: open circuit faults, short circuit faults, line-line faults, degradation faults, and partial shading.

The analysis of the abovementioned methods show that each category targets specific faults and has its own strengths. The present work presents a new IFD method combing the best assets of the three categories. In fact, it combines the use of the model from the first category, the measured I–V curve from the second category and finally the use of an intelligent bloc as described from the third category. This allows the identification of a wide range of faults. The proposed algorithm is able to detect and identify four recurrent faults: open and short circuit faults, as well as string disconnection in PV array using ANNs. Both detection and isolation are simple and fast. The developed model requires small training period and is based on only four inputs: the maximum power current and voltage from the output I–V characteristic, the solar irradiation and the cell temperature.

The remainder of this paper is organized as follow. Section 2 describes the methodology. Section 3 presents the experimental setup and data. Section 4 presents the PV sub array model. This model is used to study the residual information of the studied faults and the results are presented in Section 5. Section 6 gives a detailed

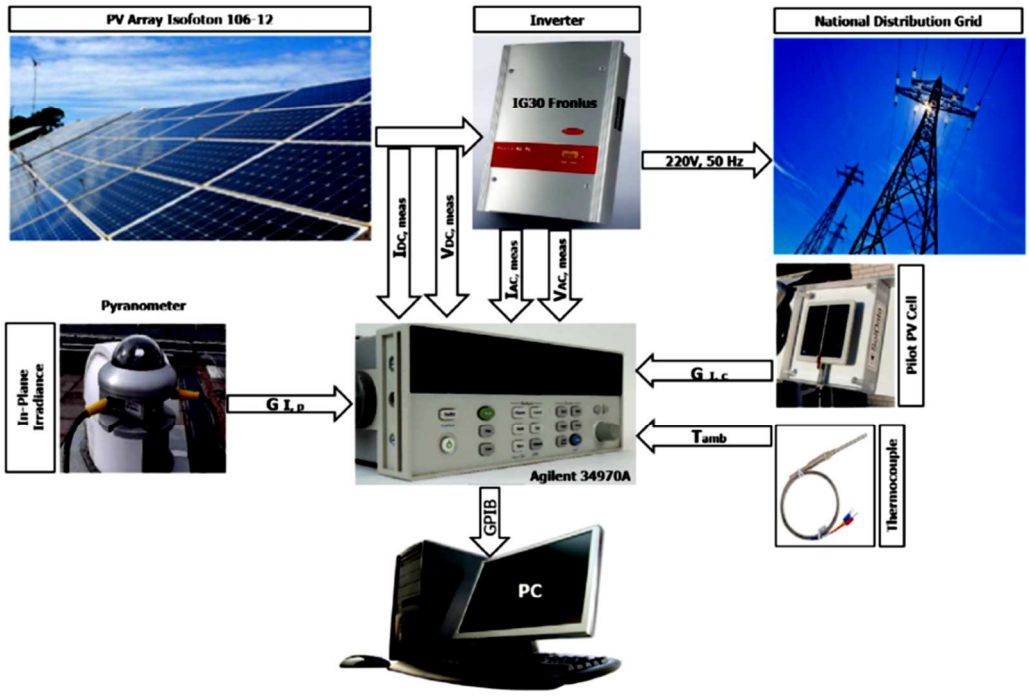
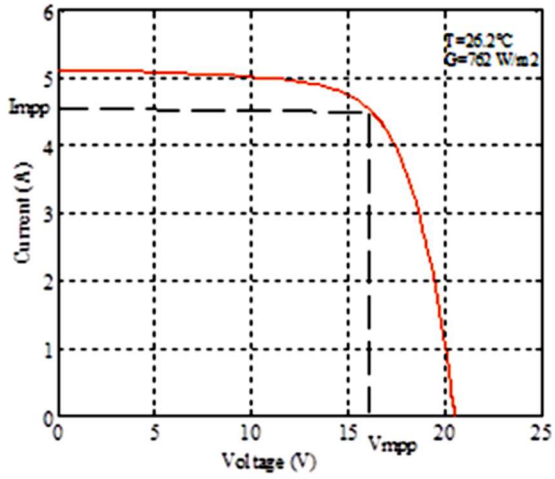


Fig. 4. Experimental setup.



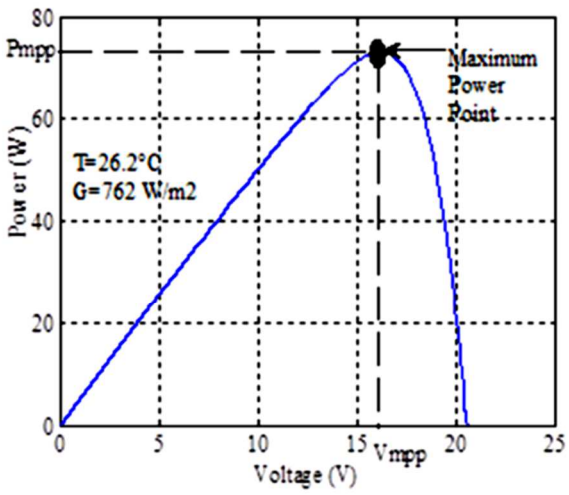
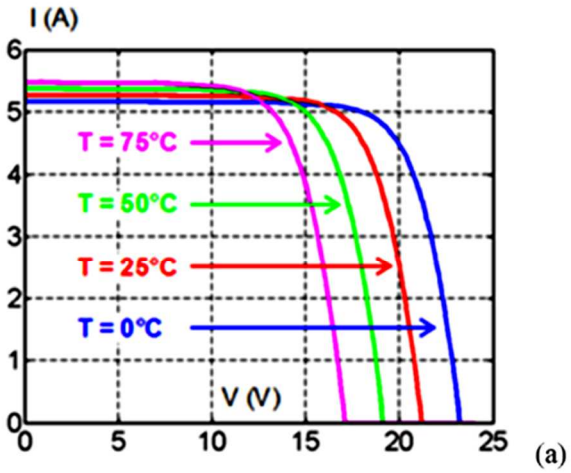


Fig. 5. I–V and P–V curves.



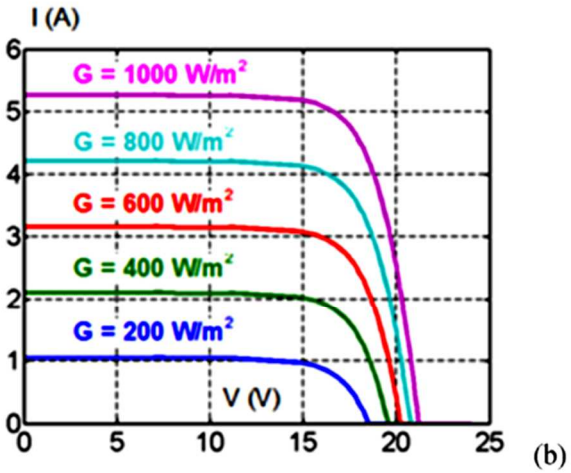


Fig. 6. Effect of (a) temperature and (b) irradiance on the I–V curve.

description of the proposed IFD algorithm while Section 7 is devoted to the presentation of the obtained results and their discussion. Finally, the conclusions and the perspectives of this approach are given in Section 8.

2. Methodology

As described by Fig. 1, the implementation of the present IFD is as a four-step procedure: 1) The detection phase: the detection of the fault is performed by calculating residual information by comparing the measurements from the system to the model outputs. This is an optimization problem which requires the calculation of the difference cost criteria. 2) The isolation phase: during this stage, two ANNs' classifiers are used to classify the information about the fault which caused the detected residual information. These two ANNs have been already trained to separately classify the current and the voltage at the MPP. 3) The identification phase: during this final stage, the output of the ANNs is analysed by a combinational algorithm called the logical block whose role is to identify and recognize the corresponding fault.

3. Experimental setup and data description

The present study uses an experimental setup located at Algiers, Algeria. This is a roof grid connected PV plant which contains, as illustrated by Fig. 2, 90 monocrystalline PV modules. The electrical properties of the PV modules are summarized in Table 1. Fig. 3 illustrates the output power of the setup and the corresponding weather conditions (solar irradiation and ambient temperature)

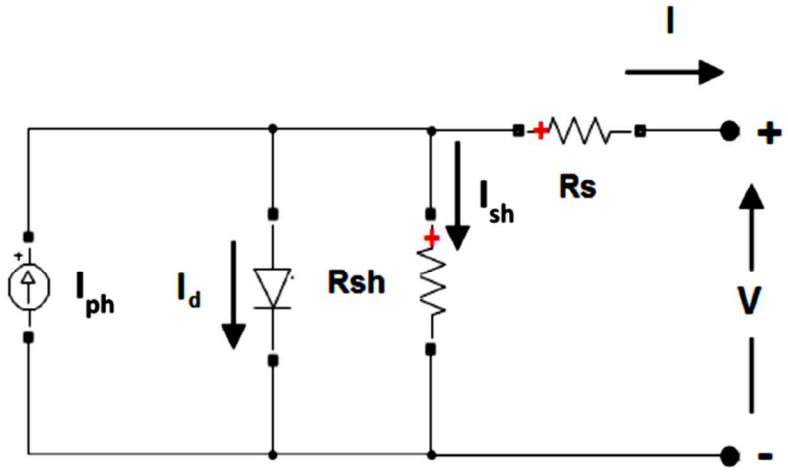


Fig. 7. One diode model of the PV module.

during one day. The experimental setup has been designed to contain four sub arrays to meet the needs of different research topics. The PV array under study is a sub array which contains 16 PV modules interconnected in two string connected in parallel, each one contains 8 modules interconnected in series.

The studied faults have been simulated on this sub array model and the corresponding data have been measured using an Agilent 34970 A data logger switch unit. The working conditions, the solar irradiation and the temperature have been measured using thermoelectric pyranometer and a K-type thermocouple respectively. Fig. 4 shows the Experimental setup.

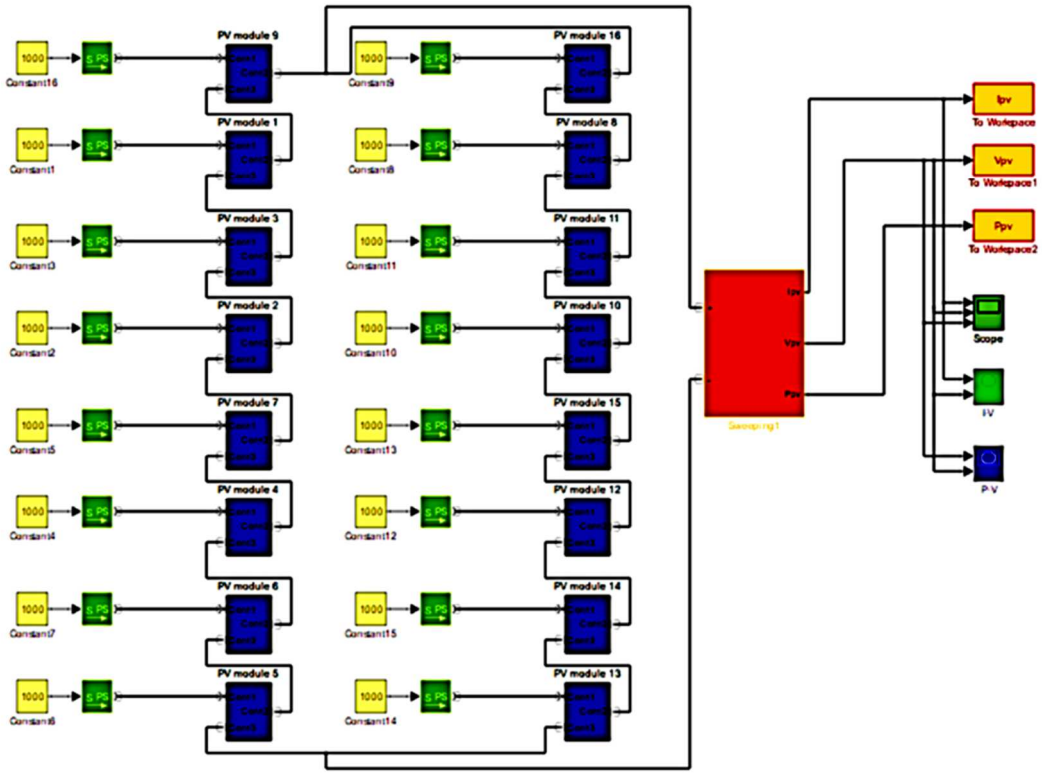
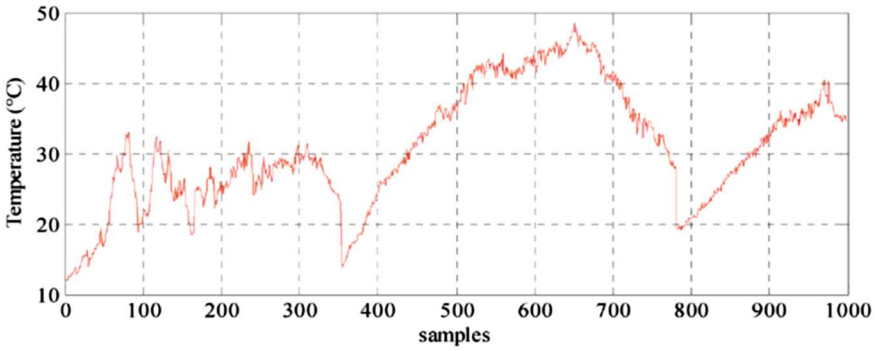
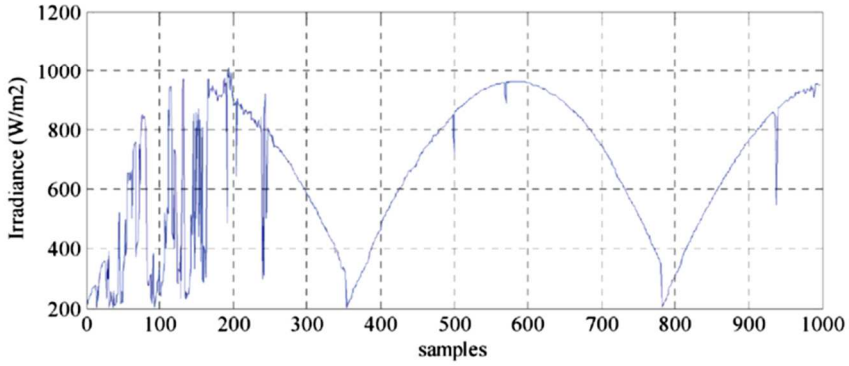


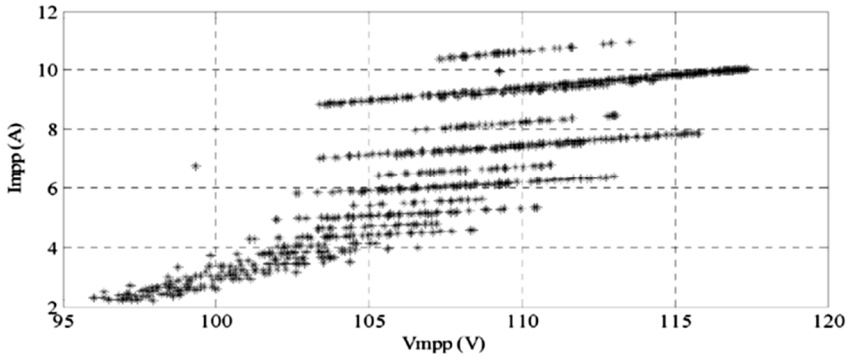
Fig. 8. Block diagram of the studied PV sub-array in Simscape.



a) The temperature



b) The irradiance



c) The current and voltage at the MPP

Fig. 9. Experimental validation data.

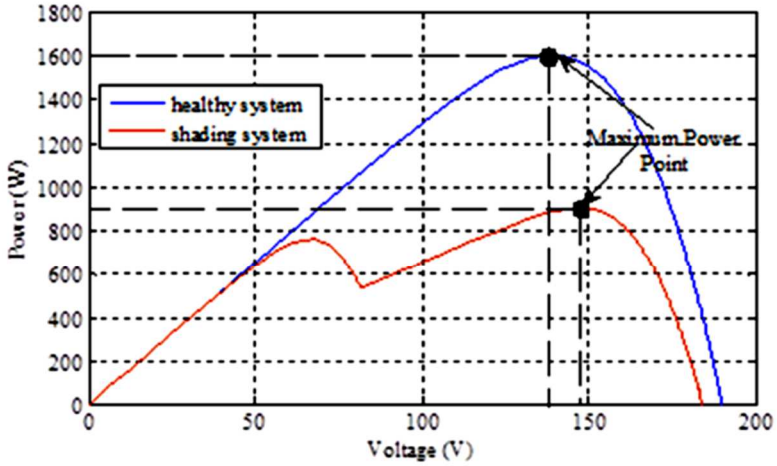


Fig. 10. Examples of temporary PV faults.

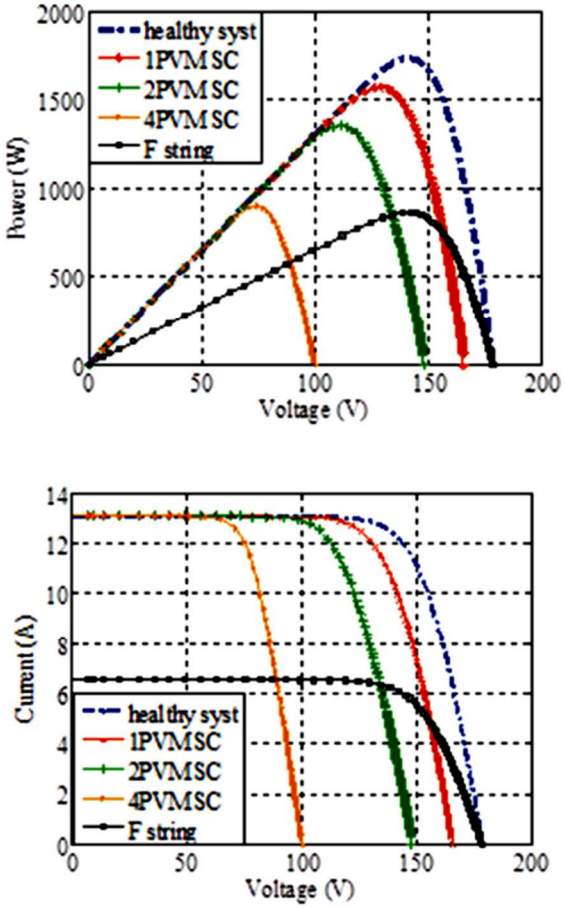


Fig. 11. Examples of permanents PV faults.

4. PV array modelling

Model based fault diagnosis need accurate models to perform fault diagnosis. Here, the mathematical model outputs of the PV array are needed to be compared with the outputs of the PV array under study. It is important to note that model based diagnosis is highly dependent on the accuracy and on the complexity of the used model. PV array modelling has been performed in three steps:

4.1. PV array modelling

Fig. 5 presents a nonlinear I–V characteristic of PV module. These characteristics are dependent on solar irradiation level and ambient temperature as illustrated by Fig. 6. These two variables are called working conditions and are the inputs of the PV module/array model while the I–V curve is its output. The I–V curve is characterized by three main points; the short circuit current I_{sc}, the open circuit voltage V_{oc} and the maximum power point (MPP) at which the power is at its maximum. The current and the voltage at the MPP are called V_{mpp} and I_{mpp} respectively.

In the literature, several models have been developed and presented to describe the behaviour of PV modules [51,52]. Here, the one diode based model described by Fig. 7 and Equation (1) is used. Where I and V are respectively the output current and voltage of the PV module, I_0 is the diode saturation current, I_{ph} is the photo generated current, R_s and R_{sh} are the series and shunt resistance respectively, T_c is the cell or module temperature, k is the Boltzmann constant ($1.3806503 \times 10^{-23}$ J/K), q is electron charge ($1.60217{,}646 \times 10^{-19}$ C) and n is the diode ideality factor. This model is called 5 parameters one diode model. The five parameters are I_0, I_{ph}, R_s, R_{sh} and n. Modelling the PV module/sub array using this model is made by finding the best combination of these 5 parameters so the output of the model can give the best fit of the experimental data.

$$I = I_{ph} - I_0\left(\exp\frac{q(V + R_s I)}{nkT_c} - 1\right) - \frac{V + IR_s}{R_{sh}} \quad (1)$$

4.2. PV array model simulation

PV array modelling has been carried out using Matlab software. Two simulation strategies are possible. The first is the simulation of equivalent circuit model by functional equations using the script language of Matlab. The second is the simulation of the equivalent circuit model blocks using Simscape. The advantage of Simscape is that it enables to rapidly create models of physical systems within the Simulink environment. With Simscape, it is possible to build physical component models based on physical connections that directly integrate with block diagrams and other modelling paradigms [53]. For this work, the second option has been chosen because it allows to simulate the targeted faults as the short circuit and the disconnection faults. Fig. 8 illustrates the interconnected 16 PV modules based sub array.

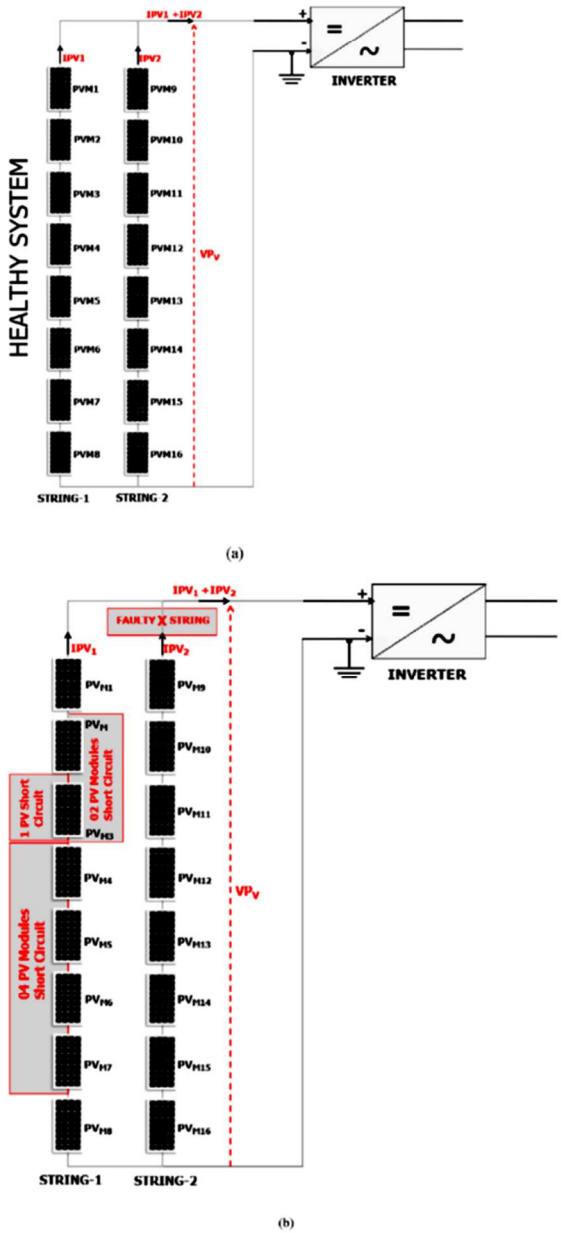


Fig. 12. Overall scheme of the PV system (a) normal operation conditions (b) faulty string and short circuits faults.

Table 2
The different state of the system with faults and their symbols.

Fault	Symbol
Normal operation conditions	C1
Fault detection refers to one panel short circuit	C2
Fault detection refers to two panels short circuit	C3
Fault detection refers to four panels short circuit	C4
Fault detection refers to string	C5

Table 3
Classifications stage, classes and symbols.

Symbols	Description	Classes
Impp n	Maximum power point current of normal operation conditions	Class1I
Impp str	Maximum power point current of faulty string	Class2I
Vmpp n	Maximum power point voltage of normal operation conditions	Class1V
Vmpp 1sc	Maximum power point voltage of one module short circuit	Class2V
Vmpp 2sc	Maximum power point voltage of two modules short circuit	Class3V
Vmpp 4sc	Maximum power point voltage of four modules short circuit	Class4V

4.3. PV array model validation

The validation of the obtained model has been done using experimental data from the experimental setup. For this, three model comparison metrics are used: root mean square error (RMSE), mean absolute percentage error (MAPE) and the coefficient of determination according to equation (2), equation (3) and equation (4) respectively [14,44]. Where y_n is the nth measured data, $\hat{y}_n$ is the nth simulated data and N is the size of the database (the number of the validation's samples). Note that the validation data are the inputs of the models which are the working condition (the solar irradiation and the ambient temperature), and the output data are the current and the voltage at the MPP as illustrated by Fig. 9.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \widehat{y_n})} \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{n=1}^{N}\frac{|y_n - \widehat{y_n}|}{|y_n|} x100\% \quad (3)$$

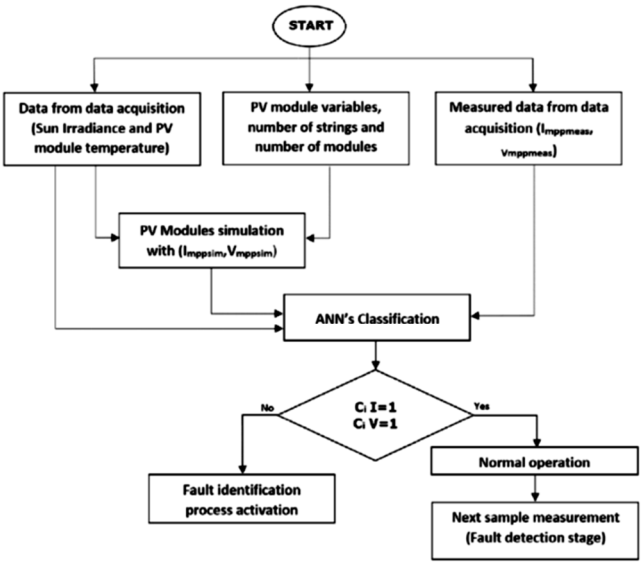
$$R^2 = 1 - \left(\frac{\sum_{n=1}^{N}(y_n - \widehat{y_n})^2}{\sum_{n=1}^{N}(y_n)^2}\right) \quad (4)$$

5. Faulty sub array versus healthy sub array

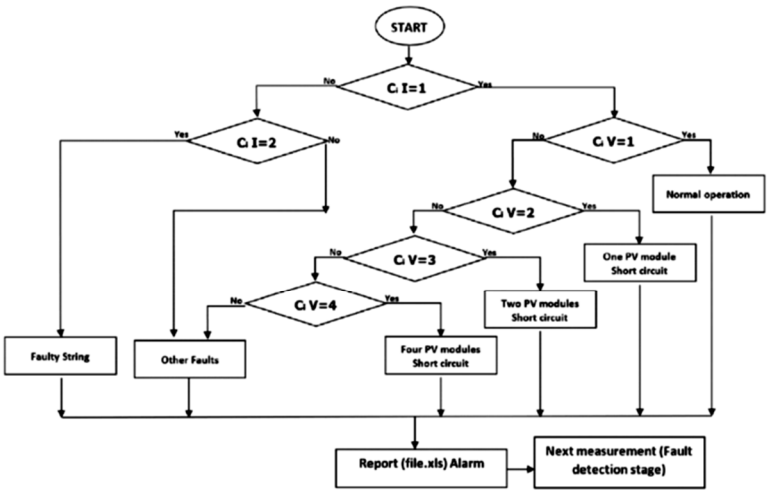
Usually, the faults occurring in a photovoltaic system are divided

Table 4
Different combination of classes obtained.

Impp	Impp Class	Vmpp	Vmpp Class	Global Description	Global Classification
Impp n	Class1I	Vmpp n	Class1V	Normal operation conditions	Class1
Impp n	Class1I	Vmpp1sc	Class2V	One faulty PV module in string	Class2
Impp n	Class1I	Vmpp2sc	Class3V	Two faulty PV modules in string	Class3
Impp n	Class1I	Vmpp4sc	Class4V	Four faulty PV modules in string	Class4
Impp str	Class2I	Vmpp n	Class1V	Faulty string	Class5



a) Fault detection stage



b) Fault identification stage

Fig. 13. Flowchart of the IFD algorithm.

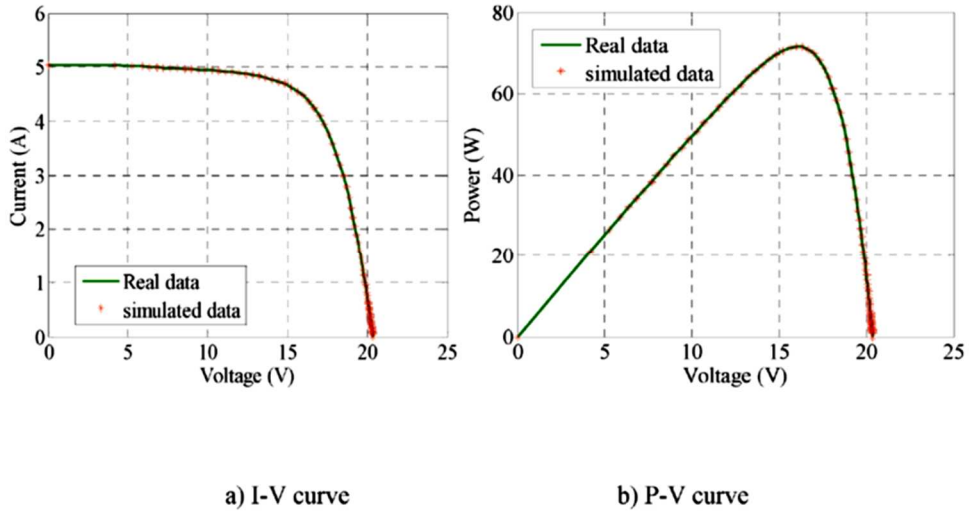


Fig. 14. Measured and simulated PV module output data.

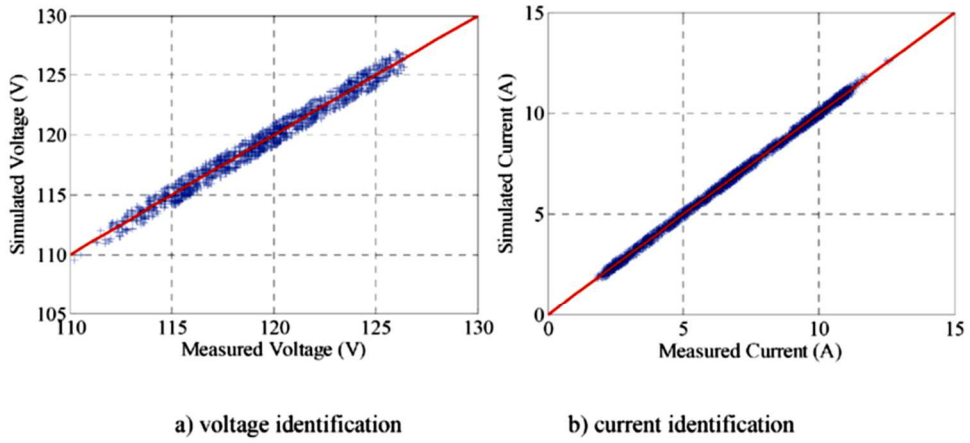


Fig. 15. Identification of the maximum power point in normal operation conditions.

Table 5
Model performance metrics values.

	RMSE	MAPE (%)	R^2
Impp	0.0433	2.06	0.9999
Vmpp	0.0606	5.23	0.9926

into two classes: temporary faults and permanent faults. Temporary faults like shading may occur due to cloudy weather, snow, dusty and sandy PV array, in addition to building, trees, leaves and bird excrement. The I–V characteristics of PV string with shading effect are illustrated by Fig. 10. In this paper, those temporary faults are not taken into consideration where the fault is automatically reset after a given period of time and the system returns back to normal operation conditions. However, permanent faults are the most crucial to be analysed notably short circuit and open circuit since they are enduring and ultimately requiring human intervention. Faults due to short circuits have an important impact on cells, bypass diodes and modules/strings. These faults can be originated by humidity invasion, loss of wiring connection between modules/strings and aging process. However, open circuit faults are caused by a sudden disconnection between cells or between modules/strings of the PV array. The I–V characteristics corresponding to these faults can cause important power losses as illustrated by Fig. 11. These changes at the I–V characteristic's shape could therefore be used to identify the corresponding faults.

6. IFD algorithm's description

The IFD algorithm is designed to detect, identify and isolate four faulty modes: 1) PV module short circuit, 2) two PV modules short circuit, 3) four PV modules short circuit and 4) faulty string. These faults are illustrated by Fig. 12 and summarized by Table 2.

The diagnosis is based on a two-stage procedure. In the first

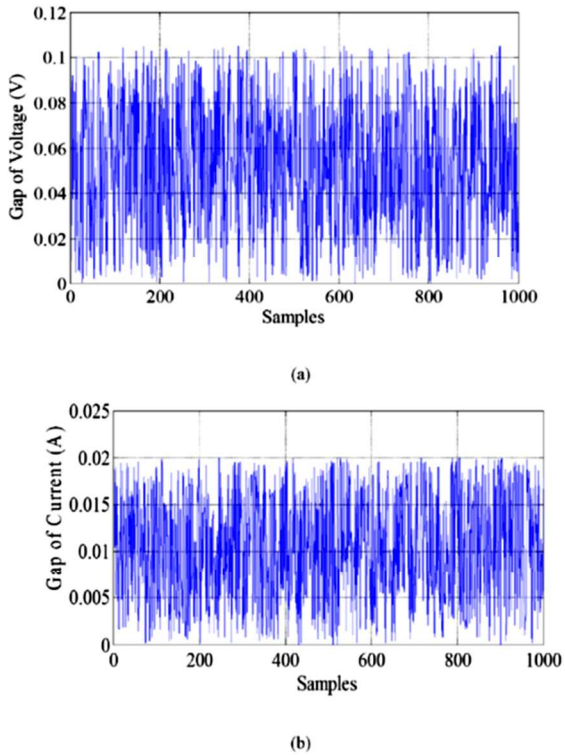


Fig. 16. Absolute error between measured and simulated voltage and current at MPP.

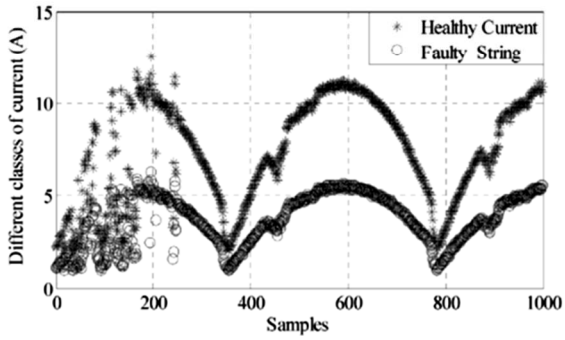


Fig. 17. The impact of the string fault on the current.

step, the fault is detected by comparing the PV system's and its mathematical model outputs. In case of fault detection, the identification stage is launched. In both stages (ANN) are used. They are exploited to simulate the PV system in the first stage and for the fault classification during the second stage. The classification is based on two ANNs, the first one (ANN-CI) is used for the Impp classification and the second's (ANN-CV) for the Vmpp classification. For the identification stage, a combinational algorithm is used to analyse the findings of the classification stage which are based on the Impp and Vmpp values (codes). The different classes and their combinations are summarized in Tables 3 and 4 respectively,

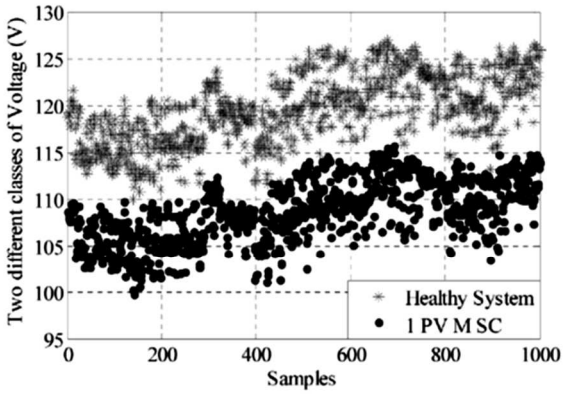


Fig. 18. The voltage of a PVG with one short-circuited module.

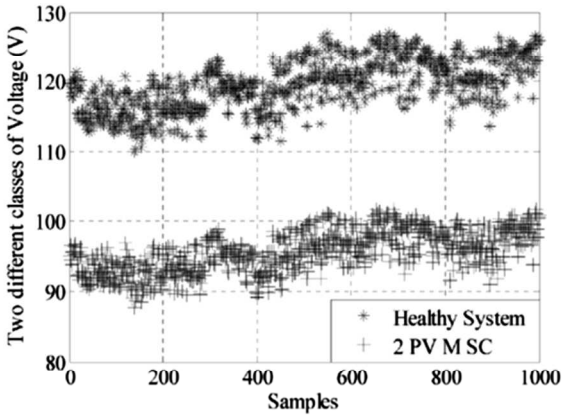


Fig. 19. The voltage of a PVG with two short-circuited modules.

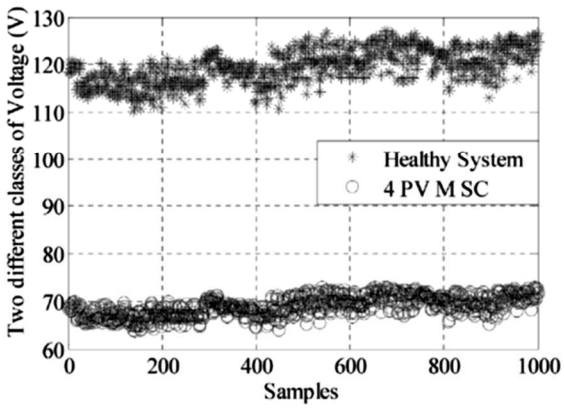


Fig. 20. The voltage of a PVG with four short-circuited modules.

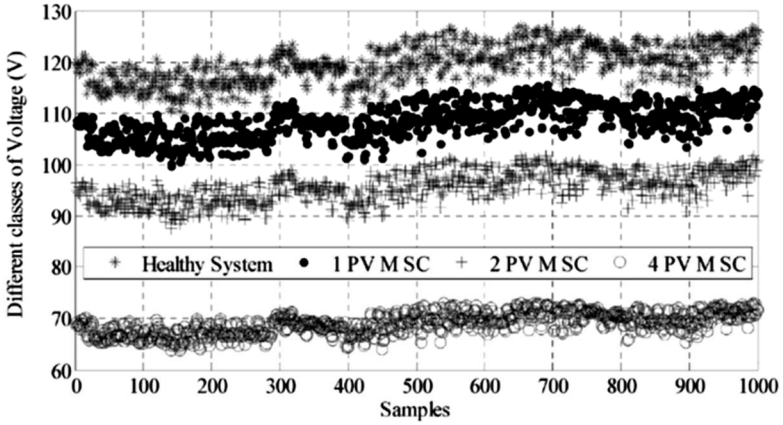


Fig. 21. The voltage of global different short-circuited modules in PVG.

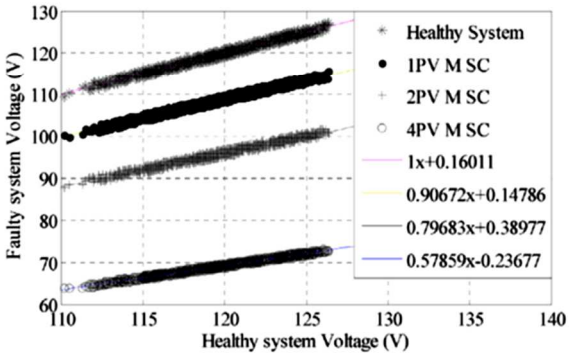


Fig. 22. Faulty simulated voltage via healthy measured voltage of the system.

while the flowchart of the IFD algorithm is illustrated in Fig. 13.

7. Test results and discussion

The effectiveness of the IFD Approach relies at once on its PV module's modelling precision and faults classification accuracy. For the first criterion, the obtained results show good agreement between the measured data and the model generated data as illustrated by Fig. 14. The PV module/generator modelling accuracy is the sole guarantee for MPP identification and thus of the faults detection. The identification results of the MPP voltage (Class1V) and current (Class1I) are shown in Fig. 15. The corresponding error metrics are summarized in Table 5. The correlation coefficient between the real and simulated curves is 0.99 for both voltage and current, which leads to an absolute error which do not exceed 0.11 V and 0.02 A for the voltage and current at MPP respectively as shown in Fig. 16.

This developed model has been used to detect the faulty

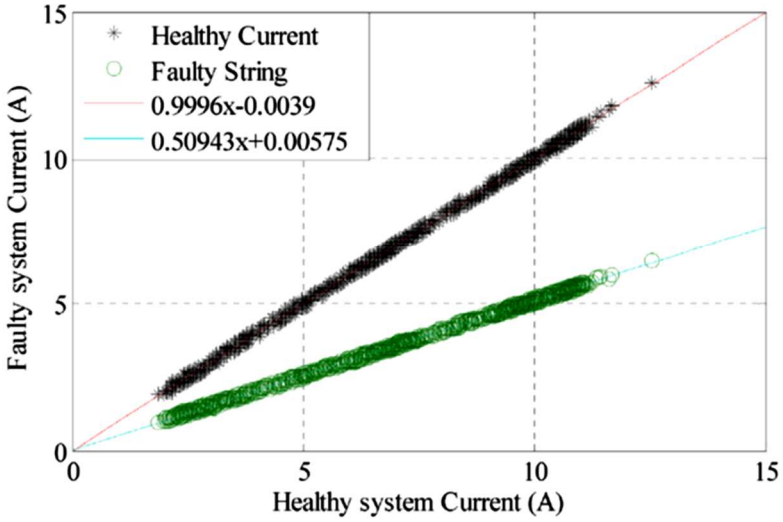


Fig. 23. Faulty simulated current via healthy measured current of the system.

behaviour of the PV system based on the deviation between the desired output and the real output under the same working conditions. This difference between the model output and the system output is the residual information. In Fig. 17, the studied data show that the string fault leads to a remarcable diminution of the current. For the 1000 analysed samples, the faulty current has been decreased by 50%. It should be noted that this remarcable diminution of the current guarantees the corresponding fault detectability and minimize any risk of confusion. However, this does not extend to the voltage residual information as shown by Fig. 18 which shows that confusion can arise in the overlap area between the two voltages in case of one short-circuited PV module. In fact while MPP voltage of the healthy PVG varies from 112 V to 127 V, the faulty PVG voltage varies between 99 V and 116 V. Fig. 19 shows that this overlap does not exist when two PV modules are short-circuited in the same string. In this case, the MPP voltage of the faulty PVG varies between in the interval 99 V and 116 V. This diminution is more remarcable when four PV modules are short-circuited in the same string. As illustrated by Fig. 20, the MPP voltage of the faulty system varies between 65 V and 74 V. This close correlation between the variation's interval of the faulty PVG voltage and the number of the short-circuited modules can be used to detect and identify the fault as illustrated by Fig. 21.

The comparison between these three faults classes and the healthy has been used to establish the correlation illustrated by Fig. 22. The PVG have a 10% lower voltage when only PV module is short-circuited, 25% and 48% lower voltages when two and four short-PV modules are short-circuited respectively. Concerning the current, the faulty system shows a 50% fall compared to the healthy system as illustrated by Fig. 23.

For the fault identification, the IFD algorithm combines the voltage residual information and the current residual information. This mixed information enables to build the following correspondence Table 6.

The end of this section consists to simulate the diagnosis of the model using two ANN's trained and tested, the output of both ANNs represents the input of combinational algorithm in order to obtain the final classification.

The obtained database presents 10,000 samples of each attribute, divided in two parts: 5000 samples for the ANN current classification and 5000 samples for the ANN voltage classification.

Both ANNs used tangent sigmoid as the activation function. In the training phase, both ANNs have been trained using Levenberg Marquart algorithm taking 2000 samples for each attribute whether electric or climatic.

The first ANN classifies the current of the maximum power point. This ANN contains two neurons in input layer, which represent irradiance and current for the maximum power point respectively, one neuron in output layer, which represent current classification, between these two layers, there are two hidden layers of eight neurons in each one. As mentioned in the paragraph above, in the training phase, each sample contains the current value at the MPP (Impp (A) and the irradiance level (W/m^2) that is (2000 × 2 = 4000 data) as input. 50% of the samples represent healthy operation conditions and remaining 50% represent the disconnected string with the combination summarized in Table 7.

Table 6
Faults correspondence table.

	Impp (%)		*Vmpp (%)*		*Pmpp (%)*	
Healthy system		100		100		100
1 PV M SC	↘	1	↘	10	↘	11
2 PV M SC	↘	1	↘	25	↘	26.5
4 PV M SC	↘	1	↘	48	↘	49
Faulty string	↘	50		100	↘	50

Table 7
Faults correspondence table.

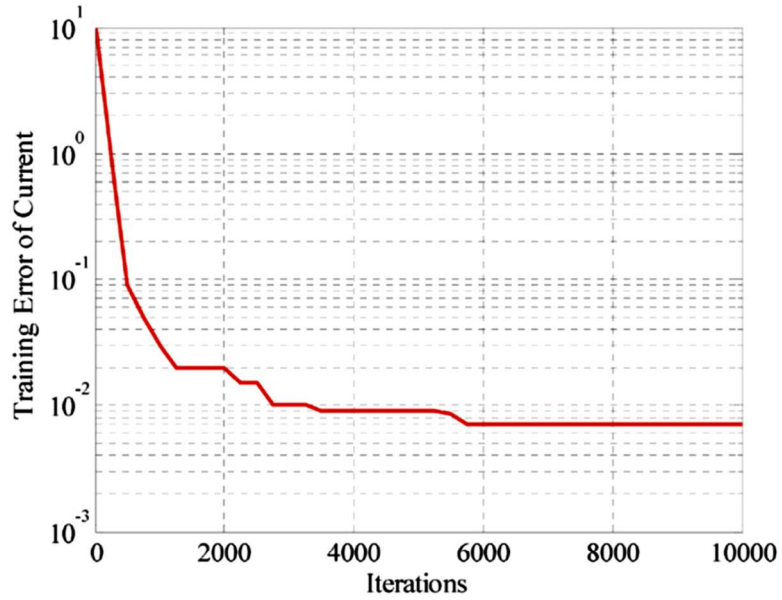
Faults	Number of simples for each attribute
Healthy current	50 samples for each attribute.
Four PV module short circuited	40 samples for each attribute.
One PV module short circuited	100 samples for each attribute.
Two PV module short circuited	52 samples for each attribute.
Healthy voltage	30 samples for each attribute.
Two PV module short circuited	70 samples for each attribute.
One PV module short circuited	20 samples for each attribute.
Four PV module short circuited	38 samples for each attribute.
Healthy current	11 samples for each attribute.
Disconnected string	58 samples for each attribute.
Healthy current	31 samples for each attribute.

The second ANN classifies the voltage of the maximum power point. This ANN contains two neurons in input layer and one neuron in the output layer. The input neurons receive cell temperature and voltage of the maximum power point respectively when the output neuron gives the voltage classification. The ANN contains two hidden layers of thirteen and twenty neurons respectively. For its training, 2000 samples have been used for each attribute (Vmpp (V), Temperature (°C)) it means (2000 × 2 = 4000 data) as input, for training phase; 25% samples represent healthy voltage (500 samples for each attribute) and the remaining 75% samples represent different faults divided into three equal categories which are: 25% for one PV module short circuited, 25% for two PV module short circuited and 25% for four PV module short circuited. For the ANN diagnosis 500 samples has been employed of each attribute (Vmpp (V), Temperature (°C)) distributed in eleven cases respectively as summarized in Table 8.

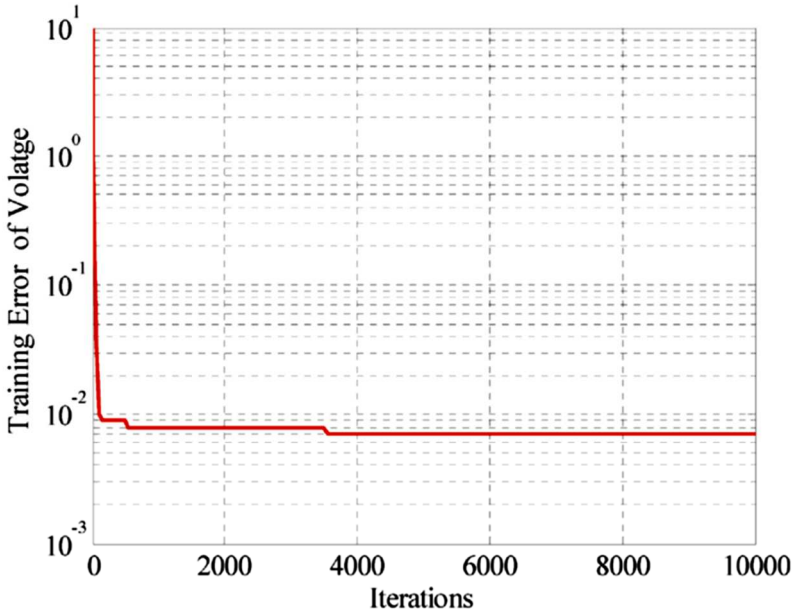
The evolution of the training error of both ANNs is shown in Fig. 24. The classification results are illustrated by Figs. 25 and 26 for the first ANN and the second ANN respectively. This shows the ability of the obtained ANNs to introduce the incoming data in their right classes with high accuracy: 99.6% for the current classification and 99% for the voltage classification. The accuracy of the current and voltage classification is very important for the whole algorithm since it is the first step of diagnosis on which depends the fault class identification, the second and final step of the diagnosis. As illustrated by Fig. 27, the majority of the output decisions are in their right classes, the few confusion cases are 1) between healthy system (C1) and one PV module short circuited (C2), and between two PV modules short circuited (C3) and one PV module short circuited (C2) and 2) two confusion cases for healthy system (C1) belonging to faulty string (C5). The analysis of these confusion cases revealed that their causes are the variation of the irradiance and the

Table 8
Faults correspondence table.

Faults	Number of simples for each attribute
Healthy voltage	50 samples for each attribute.
Four PV module short circuited	40 samples for each attribute.
One PV module short circuited	100 samples for each attribute.
Two PV module short circuited	52 samples for each attribute.
Healthy voltage	30 samples for each attribute.
Two PV module short circuited	70 samples for each attribute.
One PV module short circuited	20 samples for each attribute.
Four PV module short circuited	38 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Disconnected string	58 samples for each attribute.
Healthy voltage	31 samples for each attribute.



(a)



(b)

Fig. 24. ANN's training Error (a) of current (b) of voltage.

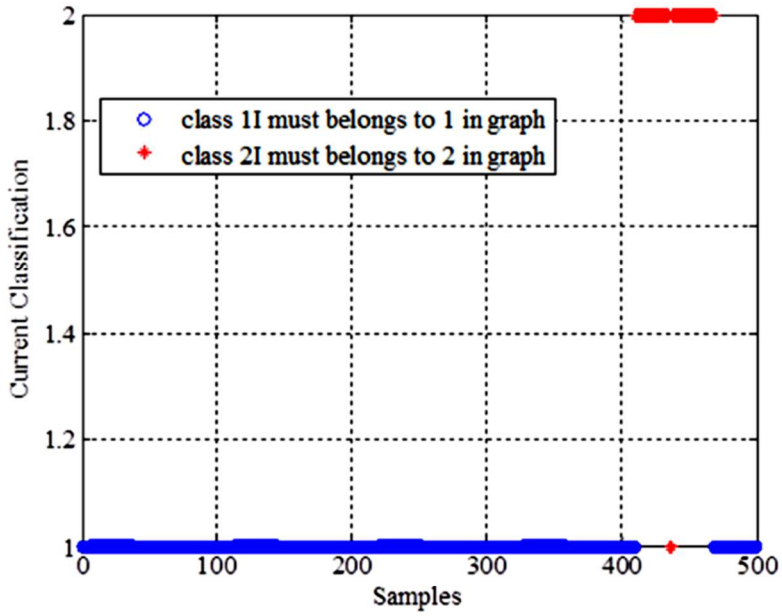


Fig. 25. Impp classification using ANN.

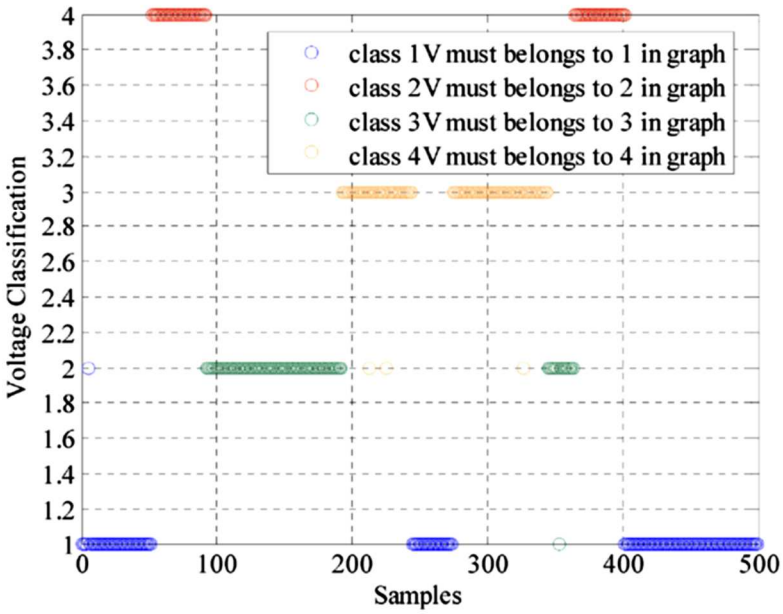


Fig. 26. Vmpp classification using ANN.

temperature respectively. However, notwithstanding these isolated cases of confusion, the final simulation results show a high performance with a good accuracy equal to 98.6%.

To show the effectiveness of the developed IFD, a test considering new samples for each attribute has been performed in order to evaluate the obtained results. For this aim 120 samples are applied for each attribute and split into seventeen cases as summarized by Table 9. The results of this test are illustrated by Fig. 28. The robustness of the IFD approach is evident since only two faulty situations among the 120 cases have been wrongly classified. The first case is due to the confusion between healthy system and one PV module short circuited. This confusion is caused by the

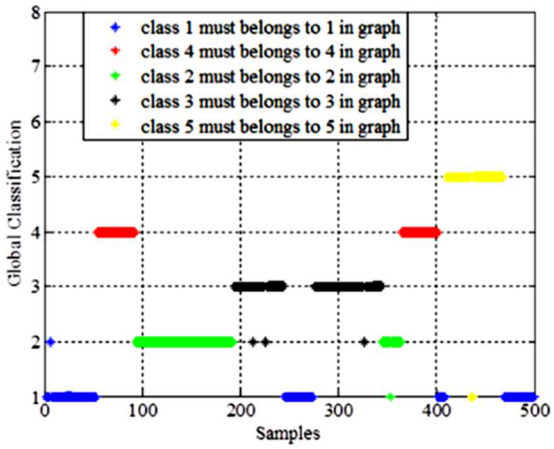


Fig. 27. Global diagnosis of the system.

Table 9
Faults correspondence table for ANN test phase.

Faults	Number of simples for each attribute
Healthy system	10 samples for each attribute.
One PV module short circuited	8 samples for each attribute.
Healthy system	4 samples for each attribute.
Disconnected string	6 samples for each attribute.
Healthy voltage	6 samples for each attribute.
Two PV module short circuited	8 samples for each attribute.
Healthy voltage	6 samples for each attribute.
Four PV module short circuited	11samples for each attribute.
Healthy system	6samples for each attribute.
Disconnected string	5 samples for each attribute.
Two PV module short circuited	8 samples for each attribute.
Healthy system	7samples for each attribute.
Four PV module short circuited	7samples for each attribute.
One PV module short circuited	4 samples for each attribute.
Healthy system	4 samples for each attribute.
Disconnected string	11 samples for each attribute.
Healthy system	9 samples for each attribute.

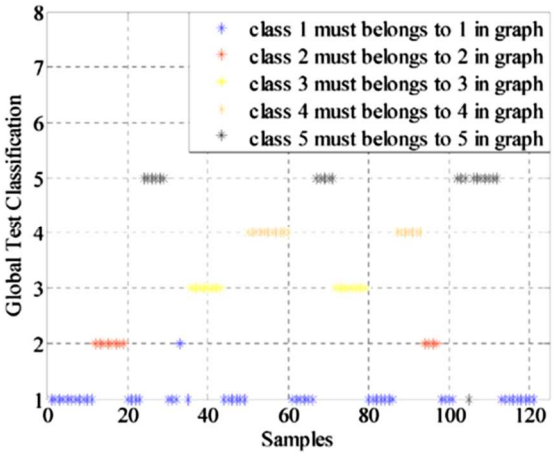


Fig. 28. IFD approach test results.

temperature variation, knowing that if the temperature increases the voltage decreases and vice versa. The second is the result of confusion between the healthy system and the disconnected string. This confusion is due to the irradiance variation, knowing that the irradiance is proportional to the current in increase or decrease. The accuracy of the IFD approach for this test is equals to 98.34%.

8. Conclusion

The present work proposes a new intelligent algorithm for PV systems' diagnosis and fault detection (IFD) for grid-connected photovoltaic systems. The fault detection and diagnosis of PV systems is necessary not only to increase system power generation reliability but also for operating costs reduction. The proposed approach is proceed in four steps: 1) the comparison of the measured data from the system to the model output based on Simscape/Simulink MATLAB, 2) the elaboration of an important database for healthy and faulty operation, 3) the classification of the faults information using two trained and finally 4) the identification and recognition of the corresponding fault.

The proposed work, which is based on measured data, guarantees four faulty operating cases: one PV module short circuited in PV string, two PV modules short circuited in PV string, four PV modules short circuited in PV string and one string modules disconnection in a PV array. For a high efficiency of the diagnosis each electrical parameter, which is considered as the fault signature, is classified separately. The final diagnostic tests were successful with matching rates equal to 99.6% and 99% for the classification stage and 98.6% for the fault identification and isolation during the final phase of the diagnosis process.

Author contribution

Chérifa Kara Mostefa Khelil: State of the art; experimental test, Programming; Results interpretation; Manuscript writing. Badia Amrouche: Diagnosis problematic explanation; Methodology conception; Paper writing and correction; General orientations. Abou soufiane Benyoucef: Programming. Kamel Kara: This work is part of his research project. The idea; The laboratory equipments; Work and results validation. Aissa Chouder: Context explanation; Experimental test and data; Results analysis; Results validation

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

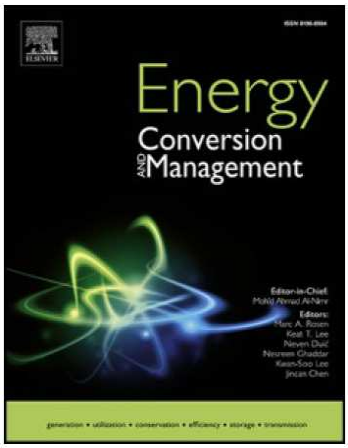
[1] Burnham L. Renewable energy: sources for fuels and electricity. Island Press; 1993. p. 1160.
[2] IEA. A snapshot of global markets 2020: Report IEA-PVPS T1-37.
[3] Harrou F, Fillatre L, Nikiforov I. Anomaly detection/detectability for a linear model with a bounded nuisance parameter. Annu Rev Contr 2014;38(1):32–4.
[4] Benyoucef A, Chouder A, Kara K, Silvestre S. Artificial bee colony based algorithm for maximum power point tracking (MPPT) for PV systems operating under partial shaded conditions. ppl. Soft Comput 2015;32:8–38.
[5] Chouder A, Silvestre S. Automatic supervision and fault detection of PV systems based on power losses analysis. Energy Convers Manag 2010;51:1929–37.
[6] Gokmen N, Karatepe E, Celik B, Silvestre. Simple diagnostic approach for determining of faulted PV modules in string based PV arrays. Sol Energy 2012;86:3364–77.
[7] Chouder A, Silvestre S. Analysis model of mismatch power losses in PV systems. Sol Energy 2009;131(2). 024504-4.
[8] Silvestre S, Chouder A, Karatepe E. Automatic fault detection in grid connected PV systems. Sol Energy 2013;94:119–27.
[9] Spataru S, Sera D, Kerekes T, Teodorescu T. Photovoltaic array condition

monitoring based on online regression of performance model. In: Proceedings of the 39th IEEE photovoltaic specialists conference; 16–21 June 2013. Tampa, Florida; 0815–20.
[10] Chine W, Mellit A, Pavan AM, Kalogirou SA. Fault detection method for grid-connected photovoltaic plants. Renew Energy 2014;66. 99–10.
[11] Platon R, Martel J, Woodruff N, Chau TY. Online fault detection in PV systems. IEEE Trans. Sustain. Energy 2015;6(4):1200–7.
[12] Drews A, de Keizer AC, Beyer HG, Lorenz E, Betcke J, van Sark WGJHM, Heydenreich W, Wiemken E, Stettler S, Toggweiler P, Bofinger S, Schneider M, Heilscher G, Heinemann D. Monitoring and remote failure detection of grid connected PV systems based on satellite observations. Sol Energy 2007;81: 548–64.
[13] Bastidas-Rodriguez JD, Franco E, Petrone G, Ramos-Paja CA, Spagnuolo G. Quantification of photovoltaic module degradation using model based indicators. Math Comput Simulat 2017;131:101–13.
[14] Dhoke A, Sharma R, Kumar Saha T. An approach for fault detection and location in solar PV systems. Sol Energy 2019;194. 197-08.
[15] Garoudja E, Harrou F, Sun Y, Kara K, Chouder A, Silvestre S. Statistical fault detection in photovoltaic systems. Sol Energy 2017;150:485–99.
[16] Chen Z, Chen Y, Wu L, Cheng S, Lin P. Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. Energy Convers Manag 2019;198:111793.
[17] Takashima T, Yamaguchi J, Ishida M. Fault detection by signal response in PV module strings. In: Proceedings of the 33rd IEEE photovoltaic specialists conference. vols. 1–5; 11–16 May 2008.
[18] Johnson J, Kuszmaul S, Bower W, Schoenwald D. Using PV module and line frequency response data to create robust arc fault detectors. In: Proceedings of the 26th European photovoltaic solar energy conference and exhibition; 05–09 September 2011. p. 3745–50. Hamburg, Germany.
[19] Takashima T, Yamaguchi J, Otani K, Oozeki T, Kato K, Ishida M. Experimental studies of fault location in PV module strings. Sol Energy Mater Sol Cell 2009;1079–82.
[20] Solórzano J, Egido MA. Automatic fault diagnosis in PV systems with distributed MPPT. Energy Convers Manag 2013;76:925–34.
[21] Sera D, Teodorescu R, Rodriguez P. Photovoltaic module diagnostics by series resistance monitoring and temperature and rated power estimation. In: Proceedings of the 34th annual conference of IEEE industrial electronics (IECON); 10–13 November 2008. p. 2195–9.
[22] Tina GP, Cosentino F, Ventura F. Monitoring and diagnostics of photovoltaic power plants. London, United Kingdom: World Renewable Energy Congress; 2014.
[23] SolarEdge. Performance of PV topologies under shaded conditions (white paper). SolarEdge; 2013.
[24] Kjær SB, Oprea O, Borup U. Adaptive sweep for PV applications. In: 26th European photovoltaic solar energy conference and exhibition; 2011. p. 3708–10. Hamburg, Germany.
[25] Swingler A. Photovoltaic string inverters and shade-tolerant maximum power point tracking: toward optimal harvest efficiency and maximum ROI (white paper). Burnaby, Canada: Schneider Electric; 2010.
[26] Spataru S, Sera D, Kerekes T, Teodorescu R. Diagnostic method for photovoltaic systems based on light I–V measurements. Sol Energy 2015;119:29–44.
[27] Hachana O, Giuseppe Marco Tina, Hemsas K. PV array Fault diagnostic technique for BIPV systems. Energy Build 2016;126:263–74.
[28] Mekki H, Mellit A, Salhi H. Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules. Simulat Model Pract Theor 2016;67:1–13.
[29] Shrikhande S, Varde P, Datta D. Prognostics and health management: methodologies & soft computing techniques. In: Current trends in reliability, availability, maintainability and afety. Springer; 2016. p. 213–27.
[30] Hare J, Shi X, Gupta S, Bazzi A. Fault diagnostics in smart micro-grids: a survey. Renew Sustain Energy Rev 2016;60:1114–24.
[31] Suganthi L, Iniyan S, Samuel A. Applications of fuzzy logic in renewable energy systems – a review. Renew Sustain Energy Rev 2015;48. 585–07.
[32] Silvestre S, daSilva M, Chouder A, Guasch D, Karatepe E. New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators. Energy Convers Manag 2014;86:2 41–49.
[33] Tadj M, Benmouiza K, Cheknane A, Silvestre S. Improving the performance of PV systems by faults detection using GISTEL approach. Energy Convers Manag 2014;80. 298–04.
[34] Zhao Y, dePalma J, Mosesian J, Lyons R, Lehman B. Line–line fault analysis and protection challenges in solar photovoltaic arrays. IEEE Trans Ind Electron 2013;60(9):3784–95.
[35] Yuchuan W, Qinli L, Yaqin S. Application of BP neural network fault diagnosis in solar Photovoltaic System. In: Proceedings of the IEEE international conference on mechatronics and automation; 2009. p. 9–12. Changchun, China.
[36] Syafaruddin S, Karatepe E, Hiyama T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In: Proceedings of the 16th international conference on intelligent system Application to power systems. Greece: ISAP); 2011. p. 1–6.
[37] Li Z, Wang Y, Zhou D, Wu C. An intelligent method for fault diagnosis in photovoltaic array. ICSC Part II CCIS 2012;327:10–6.
[38] Madeti Siva Ramakrishna, Singh SN. Modeling of PV system based on experimental data for fault detection using kNN method. Sol Energy 2018;173:139–51.
[39] Bastidas-Rodriguez JD, Petrone G, Ramos-Paja A, Spagnuolo G. Photovoltaic modules diagnostic: an overview. In: 39th IEEE annual conference on industrial electronics society. Vienna: IECON; 2013. 96-01.
[40] Ducange P, Fazzolari M, Lazzerini B, Marcelloni F. An intelligent system for detecting faults in photovoltaic fields. In: Proceedings of the 11th international conference on intelligent systems design and applications (ISDA); 22–24 November 2011. p. 1341–6. Cordoba.
[41] Bonsignore L, Davarifar M, Rabhi A, Tina GM, Elhajjaji A. Neuro-Fuzzy fault detection method for photovoltaic systems. Energy Procedia 2014;62: 431–41.
[42] Dhimish M, Holmes V, Mehrdadi B, Dales M. Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection. Renew Energy 2018;117:257–74.
[43] Karatepe E, Hiyama T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In: 2011 16th international conference on intelligent system application to power systems (ISAP). IEEE; 2011. p. 1–6.
[44] Chine W, Mellit A, Lughi V, Malek A, Sulligoi G, Massi Pavan A. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renew Energy 2016;90:501–12.
[45] Hussain M, Dhimish M, Titarenko S, Mather P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. Renew Energy 2020;155:1272–92.
[46] Belaout A, Krim F, Mellit A, Talbi B, Arabi A. Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification. Renew Energy 2018;127:548–58.
[47] Zhenghai L, Dazheng W, Liangliang T, Jinli R, Zhuming L. A heuristic diagnostic method for a PV system: triple-layered Particle Swarm optimization–back-propagation neural network. Energies 2017;10:226.
[48] Garoudja E, Chouder A, Kara K, Silvestre S. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. En Con Man 2017;151:496–513.
[49] Zhu H, Lub L, Yao J, Daia S, Hu Y. Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. Sol Energy 2018;176. 395–05.
[50] Zhao Y, Yang L, Lehman B, De Palma JF, Mosesian J, Lyons R. Decision based fault detection and classification in solar photovoltaic arrays. In: Twenty seventh annual IEEE applied power electronics conference and exposition; 2012. p. 93–9. Orlando, FL.
[51] Amrouche B, Guessoum A, Belhamel M. A simple behavioural model for solar module electric characteristics based on the first order system step response for MPPT study and comparison. Appl Energy 2012;91(1). 395-04.
[52] Amrouche B. Improvement and experimental validation of a simple behavioural model for photovoltaic modules. Sol Energy Mater Sol Cell 2014;(128): 204–14.
[53] Mathworks éèhttps://fr.mathworks.com/products/simscape.html. Modélisation et simulation de systèmes physiques multi-domaines. Consulté le jeudi 07 juin 2018.

APPENDIX 'B'

Second Publication

Kara Mostefa Khelil, C ., Amrouche, B., Kara, K., Chouder, A. (2021). The impact of the ANN's choice on PV systems diagnosis quality. Energy Conversion and Management, 240, 114278.



Source Normalized Impact Per Paper (SNIP): 2.375

SCImago Journal Rank (SJR): 2.743

Impact Factor: 9.709 (2020 Journal Citation Reports)

5 Year Impact Factor: 8.954 (2020 Journal Citation Reports)

Cite Score: 15.9

Energy Conversion and Management 240 (2021) 114278

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Letter to the Editor

The impact of the ANN's choice on PV systems diagnosis quality

ARTICLE INFO

Keywords
Photovoltaic installations
Faults diagnosis
Artificial Neural Network
Probabilistic ANNs
Generalized regression ANNs
Radial basis function ANNs

ABSTRACT

Fault diagnosis has become an indispensable part of PV installations to ensure their safety and reliability. The accuracy, rapidity, specificity, sensitivity and the precision of faults detection and isolation are the most pertinent criterions of the diagnosis quality. The present work examines the impact of the Artificial Neural Networks choice on these criterions. For this, five ANNs are studied: back-propagation ANNs (BPNN), generalized regression ANNs (GRNN), probabilistic ANNs (PNN) and two radial basis function ANNs (RBF). These ANNs are used to identify and locate the most frequently faults encountered in PV installations: short-circuit cases and open-circuit string cases in PV generator. Comparison study using the same PV installation, working conditions, data and the same diagnosis algorithm have been carried to confront the five ANNs to the same faults. Based on experimental data, the study shows that RBF ANNs affect the rate of reaction of the algorithm in presence of faults while BPNNs and GRNN present the best results from the point of view of its speed and its important high precision with good classification efficiency. In the other hand, the PNN marks its importance by its best results displaying 100% for all key statistical concepts comparing to the other algorithms.

1. Introduction

Nowadays, modern civilization is looking for a profound and global energy change throughout the world, from fossil fuels such as: natural gas, oil, lignite and coal resources to renewable energies. The goal of this change is to avoid catastrophic climate change that affects the health's deterioration of the current and future human generations as well as for countless other species. However, reliance on renewable energies has become an inescapable trend as it plays a key role in decarbonizing the global energy system for decades to come. In addition, the production of renewable energy exceptionally photovoltaic energy growing up in staggering rate for its cleanliness, safety, quiet, reduction in energy bills and low maintenance requirement, it will be the principal source of energy in the world in the coming years increasing from 330 TWh in 2019 to almost 3300 TWh in 2030, according to reference [1].

In order to achieve a maximum harvest of energy supplied by PV systems, certain factors must be ensured such as the best irradiance level, low PV cell temperature, a good PV panels' fixation and orientation through sunlight and avoiding the shaded area. The PV modules connections and the cleanliness of PV panels contribute significantly in energy production knowing that dirt, snow and sand hide the solar irradiance and consequently reduce the reliability and the efficiency of the whole PV installation. For this purpose, regular maintenance and diagnosis become more and more crucial.

Diagnosis has become an indispensable and crucial tool to maximize power production, reliability, efficiency, safety and quality in global PV systems PV systems. That explains why real-time fault detection and identification is attracting researchers worldwide as demonstrated by the large number of studies and investigations in the field [2–21]. The analysis of the developed techniques has made it possible to identify three different categories. The first category includes model based diagnosis methods. The basic idea of this kind of diagnosis techniques is the use of the difference between the real PV system' outputs and those of its model to detect and then isolate the faults when they happen. This difference, called the residual information, provides information and quantifies the mismatch between the PV generator and its model. In all study, the PV system and its model have the same inputs, the working conditions and/or the electrical parameters of the PV system [2–9,22–25]. However, for the outputs, the focus differs since it depends on the studied faults.

The second category treats the classification of PV systems in accordance with signal processing procedure. This includes all methods based on mathematical or statistical analysis [10,26–29]. In general, this type of diagnosis focus on measured output data insinuating the current versus voltage (I–V) curve which can bring key information about numerous faults [11,12,30,31]. Right now, the new series of PV modules containing integrated converters are completely or fractionally able to sweep the IV curve of the photovoltaic generator [32–34] that makes them very well suited to this kind of diagnosis methods.

In the third category, are classified all diagnosis methods that integrate computational intelligence based techniques including Fuzzy Logic (FL) [14,17,35–38], Particle Swarm Optimization [18], k-Nearest Neighbours (kNN) [15], Decision Tree [39] and Artificial Neural Networks (ANNs) [16,19,20,38,40–45]. The literature presents an impressive number of diagnostic techniques which exploit ANNs possibilities and advantages such as their learning and generalization abilities, the rapidity of their convergence speed without forgetting the major point representing their rigor in classification [16,18–20,30,41,44–47].

In a previous work [21], a new Intelligent Fault Diagnosis (IFD) Approach for grid connected PV systems has been developed and presented. Excellent diagnosis results have been obtained. For this, two back-propagation ANNs have been used to detect then to identify the open circuit fault, one PV module short-circuit, two PV modules short-circuit and four PV modules short circuit. Since then, another question

https://doi.org/10.1016/j.enconman.2021.114278
Received 26 February 2021; Accepted 10 May 2021

has arisen; what is the impact of the choice of ANNs on this diagnosis approach?

The present work analyses the impact of the ANNs choice on the diagnosis quality. For this, five different types of ANNs are substituted in the same IFD algorithms, and their performances are analysed and compared to provide a well-argued response to the previous question. The five neural networks have the same four inputs: solar irradiation, cell temperature, the current and voltage of the maximum power point of the I-V characteristic corresponding to the first two inputs (the working conditions).

The remainder of this article is organized into four parts. Section 2 gives a preview on the IFD algorithm presentation. Section 3 exposes the details of the comparative study and the methodology. Section 4 presents the results and their discussion. At the end, the conclusions and recommendations are set out in Section 5.

2. IFD algorithm presentation

To detect and identify faults occurring in grid connected PV system, an approach to faults classification has been developed and adopted. As expanded in [21], this solution is based on two ANNs which after being trained using a rich experimental database, they became able to guarantee the detection and the identification of three recurrent cases between healthy, three different situations of short circuit modules in PV string in addition to string disconnection in PV array. To ensure this classification, the first ANN requires two input data the current of maximum power point Impp of the PV generator as well as the irradiation level, while the second ANN also needs two input data the voltage at maximum power point Vmpp of the PV generator and the ambient temperature. The ANNs outputs, which are the classification results, are then uses as inputs to a combinational block whose role is to combine the ANNs outputs to identify the corresponding faults as described by Fig. 1. The global diagnosis of the studied PV system was able to prove its quality by its excellent accuracy and efficiency from the point of view of fault localization. Table 1 lists the classification results and shows a good global performance with an average overall accuracy of 98.6%.

3. The comparative study

This comparative study is structured in two phases. A first phase aimed at choosing the ANNs to be tested while the second part concerns their effectiveness evaluation.

3.1. Phase 1: ANNs choice

ANNs have been used with succeed in classification, pattern recognition, fault detection and diagnosis. To study the impact of the ANNs type of the performance of the IFD algorithm, several ANNs have been considered by maintaining the whole algorithm topology and by using the same learning and testing data and conditions. The five ANNs chosen for this comparative study are 1) Back Propagation Neural Network (BPNN), 2) Statistical ANNs which are considered to be the most important statistical neural networks in the literature. In the present study, two RBF ANNs are used and noted RBF1 and RBF2. The fourth ANN is a Probabilistic neural network (PNN) and the fifth ANN is a

Table 1
Performance of classification with ANN (%).

	Decision Class					
Input Class		C 1	C 2	C 3	C 4	C 5
	C 1	*99.18*	0.82	0	0	0
	C 2	0.83	*99.16*	0	0	0
	C 3	0	4.77	*95.23*	0	0
	C 4	0	0	0	*100*	0
	C 5	0	0	0	0	*99.6*

Overall accuracy = 98.6%.

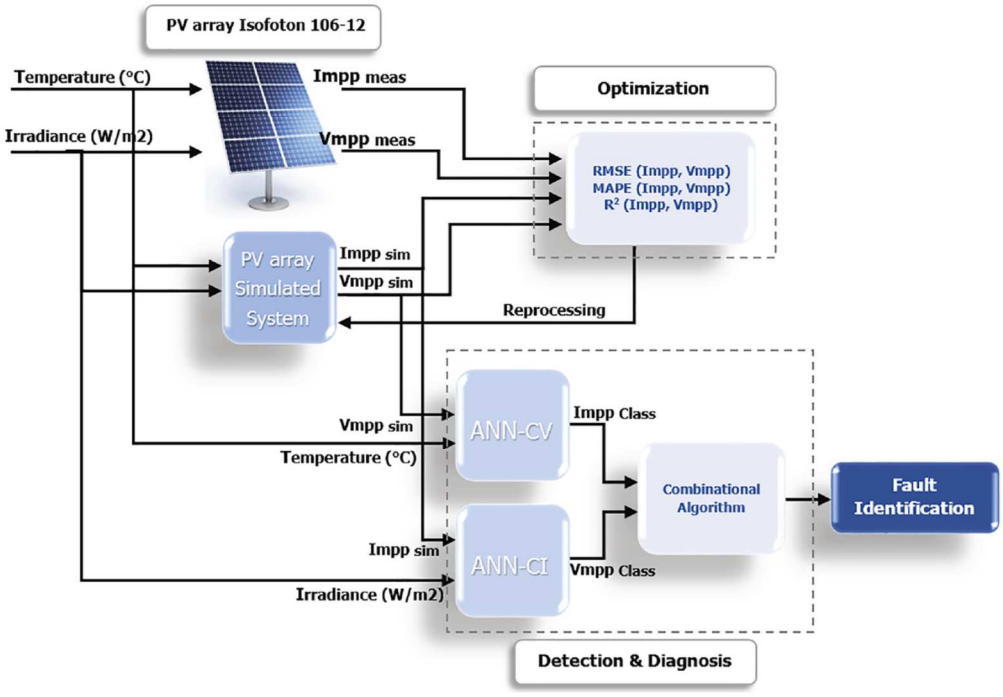


Fig. 1. Schematic description of fault diagnosis methodology [21].

Generalized Regression neural network (GRNN) both belonging to the previous family of statistical ANNs.

3.1.1. BPNN

The BPNN is an multilayer perceptron (MLP) that contains three basic parts: the input layer, at least one hidden layer and the output layer as described by Fig. 2. The middle layer is the dominant part in the network which implements a non-linear transformation to the inputs of the network. This non-linear transformation includes mainly weighting, adding a bias and applying a nonlinear limiting function such as a sigmoidal function. This transformation is described by (Eqs. (1) and (2)). With xi is the ith input of the ANN, wi is its weight, and b its bias.

$$v = \sum_{i=0}^{n} x_i w_i + b_i \tag{1}$$

$$\varphi_i(x) = \frac{1}{1+\exp(-v)} \tag{2}$$

3.1.2. Statistical neural network

This type of neural network uses statistical methods and probability theory to compare a number of random to obtain the exact decision.

3.1.2.1. RBF. Radial Basis Function Networks (RBF) are the most popular ANNs belonging to this family due to their capabilities and advantages which include better approximation capabilities, simple network structures and faster learning algorithms that make them advantageous in classification mode for nonlinear data [48]. As classical neural network, three layers compose RBF network as illustrated by Fig. 3. The particularity of these ANNs is their activation function, which is a Gaussian function as described by (Eq.(3)) [49,50]

$$\varphi_i(x) = exp\left(-\frac{\|x-c_i\|}{\sigma_i^2}\right) \tag{3}$$

where x represents the meteorological parameters, ci and σi are the center and the spread of the ith RBF node respectively. The spread σ of node coming from the intermediate layer is calculated by (Eq.(4)) [49].

With n the number of the node in the intermediate layer, *dmax* is th maximum distance between the selected cluster centers. The outputs o the non-linear activation $\varphi i(x)$ are integrated linearly with the weigh vector ωi of the output layer to produce the network output class ac cording to (Eq.(5)).

$$\sigma = \frac{d_{max}}{\sqrt{2n}} \tag{4}$$

$$class\ m = \sum_{i=0}^{n} \varphi_i w_i \tag{5}$$

In the present study two radial basis ANNs are considered. RBF1 which uses *newrbe* as activation function to create a network with zer error on training vector [51], and RBF2 which uses *newrb* activatio function. The particularity of this RBF ANN is the size of the spread which is dependent of the number of neurons and influences the qualit and generalization of changing function [51].

3.1.2.2. PNN. According to the literature, the probabilistic neura network (PNN) designed by Specht [52] contains four layers: the inpu layer, the pattern layer, the summation layer and finally the output laye displays the final classification a described by Fig. 4 [53]. The PN performance is guaranteed by two predominant factors, the number o neurons in pattern layer and the suitable activation function given b (Eq.(6)) [46]. Where wi represents the weight of the information comin to the pattern unit from the input units xi and σ represents the smoothin parameters (spread) and depends on the input data.

$$\varphi(x) = \exp\left(-\frac{(w_i-x)^t(w_i-x)}{(2\sigma^2)}\right) \tag{6}$$

The output units of the pattern layer is transferred to the units o summation layer which calculate their probability density functio (PDF) according to (Eq. (7)) [19,52,53]. Where P is the number of pat terns, n is learning set size, x_{ai} describes the corresponding i_{th} trainin pattern of a class.

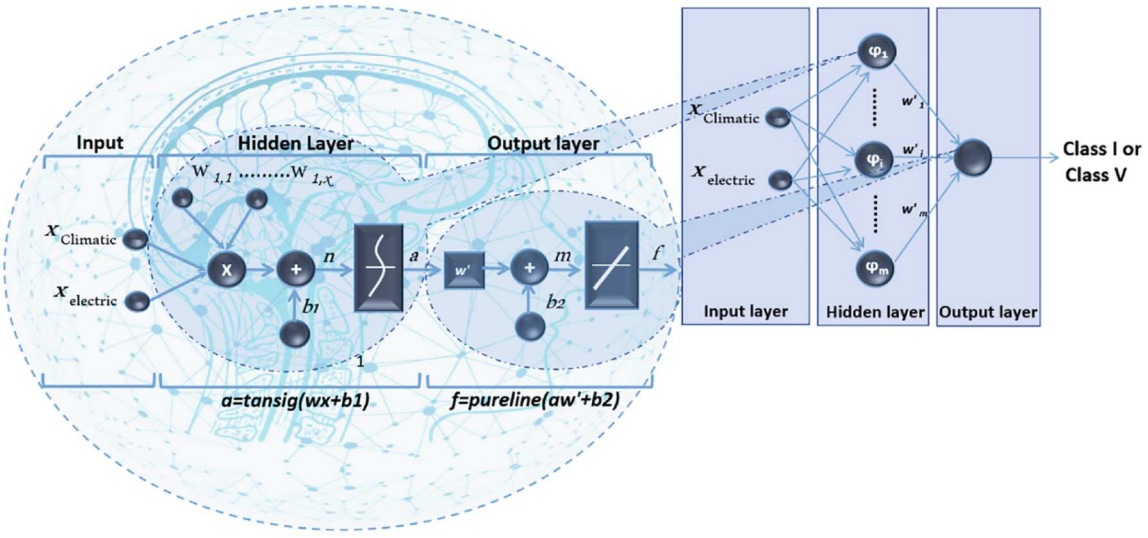


Fig. 2. BPNN architecture.

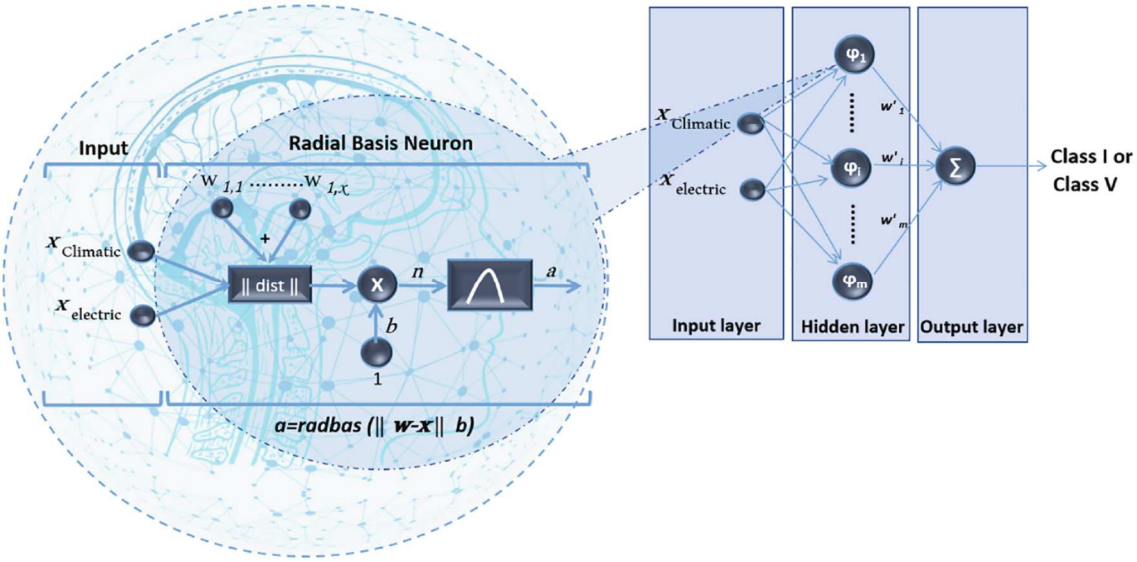


Fig. 3. RBF ANN architecture.

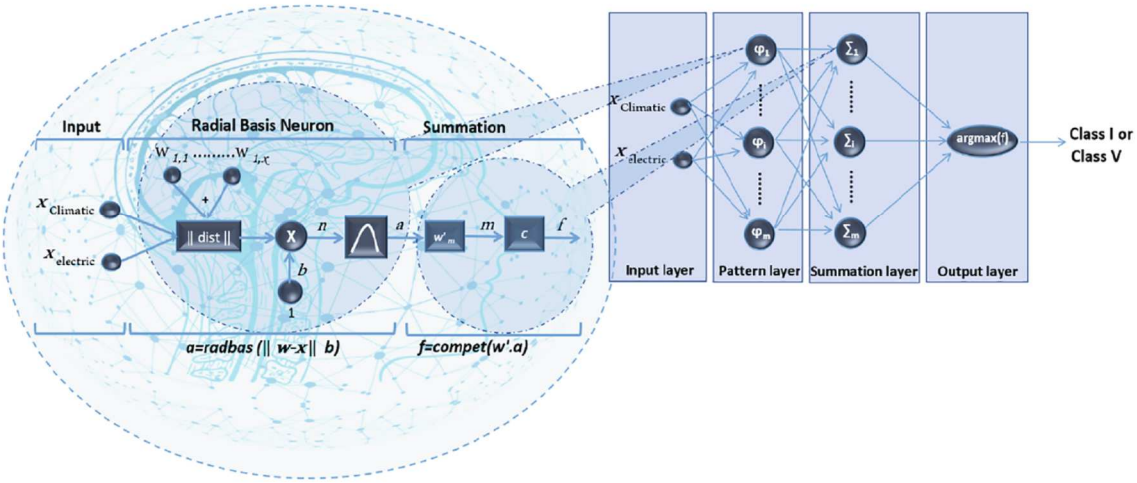


Fig. 4. PNN architecture.

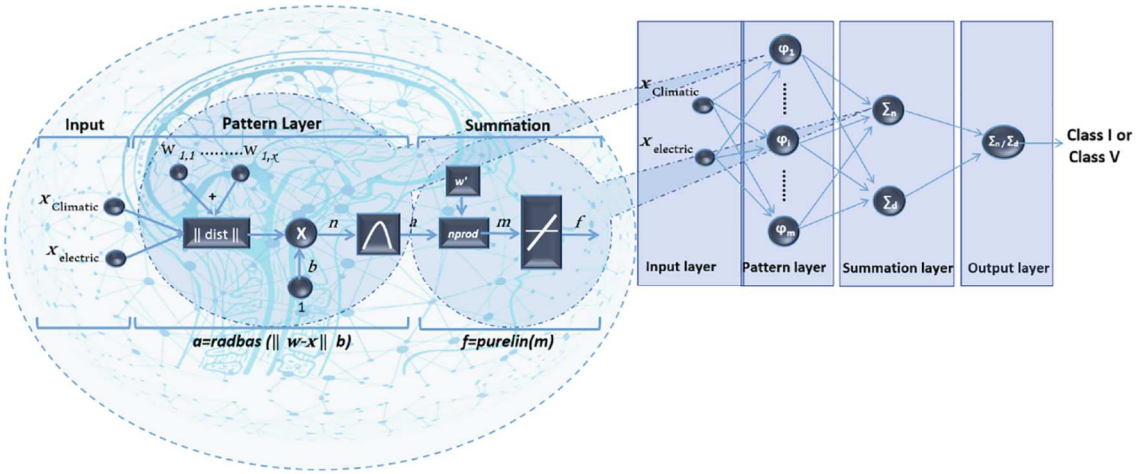


Fig 5. GRNN architecture.

$$f_a(x) = \frac{1}{(2\pi\sigma^2)^{M/2}} \frac{1}{n} \sum_{i=1}^{n} exp\left[-\frac{(w_i - x_{ai})^t (w_i - x_{ai})}{2\sigma^2} \right] \quad (7)$$

The output layer called decision layer contains one unit, which decides and displays the final class coming from summation layer as follow:

$$class(x) = argmax\{f_a(x)\}, \ 1 \le i \le M \quad (8)$$

where: M is number of classes, *class(x)*: denotes the predicted class of x.

3.1.2.3. GRNN. Like probabilistic neural networks (PNN), generalized regression neural networks (GRNNs) are recognized by their quick capability in training phase on spare data set [54]. They are organized in four layers: the input layer, the pattern layer, the summation layer and the output layer as shown in Fig. 5. The input layer receives the information data. The pattern layer plays the same role as pattern layer in PNN, its equation is described by (Eq.(9)) [54,55].

$$f_a(x, y) = \frac{1}{(2\pi\sigma^2)^{M/2}} \frac{1}{n} \sum_{i=1}^{n} \left\{ exp\left[-\frac{(X - X_i)^t (X - X_i)}{2\sigma^2} \right] exp\left[-\frac{Y - Y_i}{2\sigma^2} \right] \right\} \quad (9)$$

Where n is the number of simple observations and M is the dimension of the variable vector x, σ is a smoothing parameter, X is a particular measured value of the random variable x which represents the independent data in the input vector. The regression performed by GRNN products the most probable scalar Y provided from specified input vector x, which represents the dependent data in the output vector. Y_i is the desired scalar output corresponding to the observed input X_i. The summation layer has two kinds of processing units, the *Numerator* neuron *(N) and denominator* neuron *(D)* which are described by (Eqs. (10) and (11)) respectively. At the end, the sole neuron of the output layer receives, at its inputs, the two outputs from the summation units and divides the Numerator part by the Denominator to produce its output and then displays the classification results which the final electrical parameter classification for the present classification study $\hat{C}(x)$ according to (Eq. (12)).

Table 2
Confusion matrix under intermittent classification troubles.

Classification outcome from ANNs		Classification outcome from experimental data real label	
		True class	False class
predicted label	True class	TP	FP
	False class	FN	TN

$$Numerator(N) = \sum_{i=1}^{n} Y_i exp\left(-\frac{(X - X_i)^t (X - X_i)}{2\sigma^2} \right) \quad (10)$$

$$Denominator(D) = \sum_{i=1}^{n} exp\left(-\frac{(X - X_i)^t (X - X_i)}{2\sigma^2} \right) \quad (11)$$

$$\hat{C}(x) = Numerator / Denominator \quad (12)$$

3.2. Phase 2: effectiveness evaluation

In order to evaluate the effectiveness of the five IFD algorithms, theirs results are analyzed using the most frequently used diagnosis performance indicators in science and engineering fields [56]:

3.2.1 Accuracy: implies how nearest is the results to the real value.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (13)$$

3.2.2 Sensitivity: measures in what way the positive simples are correctly classified.

$$Sensivity = \frac{TP}{TP + FN} \times 100 \quad (14)$$

3.2.3 Specificity: measures in what way the negative simples are correctly classified.

$$Specificity = \frac{TN}{FP + TN} \times 100 \quad (15)$$

3.2.4 Precision: implies how nearest the results are to each other.

$$Precision = \frac{TP}{FP + TP} \times 100 \quad (16)$$

where:

TP: true positive, signifies that the samples contains characteristics of a specific class and indeed they are classified in this class.
TN: true negative, signifies that the samples does not contain characteristics of a specific class and indeed they are not classified in this class.
FP: false positive, signifies that the samples does not contain characteristics of a specific class and they are classified in this class.
FN: false negative, signifies that the samples contains characteristics of a specific class and indeed they are not classified in this class.

Table 2 summarizes the four major categories as result of binary classification containing two rows and two columns into confusion matrix called confusion table in the intension to confirm the performance evaluation related to the classifier. The number of rows and columns depends on the number of classes. The terms true and false refer to whether the prediction corresponds to the external criticism conversely to the terms positive and negative that refer to the prediction of the classifiers.

4. Results presentation and discussions

The IFD algorithm is based on two ANNs in addition to a combinational block whose role is to give the final classification. Therefore, to compare the effectiveness of the five chosen ANNs, two ANNs are trained for each IFD algorithm. The training database contains 12,544 samples of each attribute. It is divided in two parts: 6272 samples for the first ANN training and 6272 samples for the second ANN training task. The first ANN classifies the current at the maximum power point. This ANN contains two neurons in the input layer. Its inputs are the irradiance level and the current at the maximum power point. The output layer contains only one neuron. It gives current classification, between these two layers, there are two hidden layers of eight neurons each. The training is achieved using 50% of healthy samples and the remaining 50% represent the disconnected string. The neural current classification (diagnosis) needs 336 samples of each attribute (Irradiance (W/m2) and Impp (A)) divided into thirty-one cases respectively for testing and ensuring the classification stability and efficiency. Table 3 summarizes all treated cases with their number of samples.

The second ANN classifies the voltage at the maximum power point. This ANN contains two neurons in input layer and one neuron in the output layer. The input neurons receive cell temperature and voltage at the maximum power point while the output neuron gives the voltage classification. The ANN contains one hidden layer of forty five neurons. For its training, 2800 samples have been used for each attribute (Vmpp (V), Temperature (°C)) it means (2800 × 2 = 5600 data) as input, for training phase; 25% samples represent healthy voltage and the remaining 75% samples represent different faults divided into three

Table 3
Faults correspondence table of current.

Faults	Number of simples for each attribute
Healthy voltage	9 samples for each attribute.
Disconnected string	11 samples for each attribute.
Four PV module short circuited	15 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Two PV modules short circuited	30 samples for each attribute.
One PV module short circuited	16 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Disconnected string	9 samples for each attribute.
Two PV modules short circuited	11 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
One PV module short circuited	6 samples for each attribute.
Healthy voltage	10 samples for each attribute.
One PV module short circuited	8 samples for each attribute.
Two PV modules short circuited	7 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
Healthy voltage	4 samples for each attribute.
Disconnected string	13 samples for each attribute.
One PV module short circuited	11 samples for each attribute.
Two PV modules short circuited	14 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Disconnected string	13 samples for each attribute.
Healthy voltage	5 samples for each attribute.
One PV module short circuited	9 samples for each attribute.
Four PV modules short circuited	12 samples for each attribute.
Disconnected string	12 samples for each attribute.
Healthy voltage	21 samples for each attribute.
One PV module short circuited	13 samples for each attribute.
Four PV modules short circuited	15 samples for each attribute.
Two PV modules short circuited	10 samples for each attribute.
Healthy voltage	4 samples for each attribute.

Table 4
Voltage faults correspondence.

Faults	Number of simples for each attribute
Healthy voltage	9 samples for each attribute.
Disconnected string	11 samples for each attribute.
Four PV module short circuited	15 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Two PV modules short circuited	30 samples for each attribute.
One PV module short circuited	16 samples for each attribute.
Healthy voltage	11 samples for each attribute.
Disconnected string	9 samples for each attribute.
Two PV modules short circuited	11 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
One PV module short circuited	6 samples for each attribute.
Healthy voltage	10 samples for each attribute.
One PV module short circuited	8 samples for each attribute.
Two PV modules short circuited	7 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Four PV modules short circuited	7 samples for each attribute.
Healthy voltage	4 samples for each attribute.
Disconnected string	13 samples for each attribute.
One PV module short circuited	11 samples for each attribute.
Two PV modules short circuited	14 samples for each attribute.
Healthy voltage	5 samples for each attribute.
Disconnected string	13 samples for each attribute.
Healthy voltage	5 samples for each attribute.
One PV module short circuited	9 samples for each attribute.
Four PV modules short circuited	12 samples for each attribute.
Disconnected string	12 samples for each attribute.
Healthy voltage	21 samples for each attribute.
One PV module short circuited	13 samples for each attribute.
Four PV modules short circuited	15 samples for each attribute.
Two PV modules short circuited	10 samples for each attribute.
Healthy voltage	4 samples for each attribute.

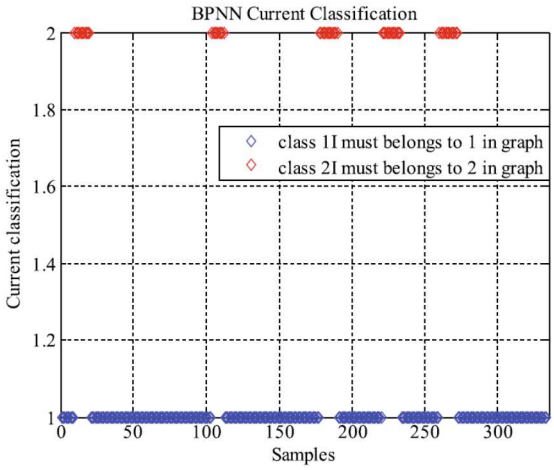


Fig. 6. Impp classification using BPNN.

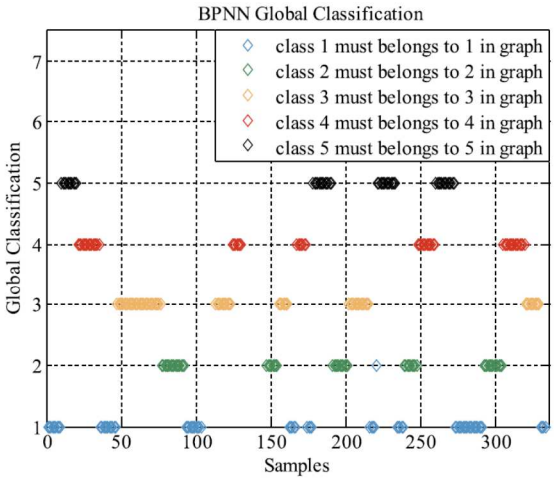


Fig. 8. Global diagnosis of the system using BPNN.

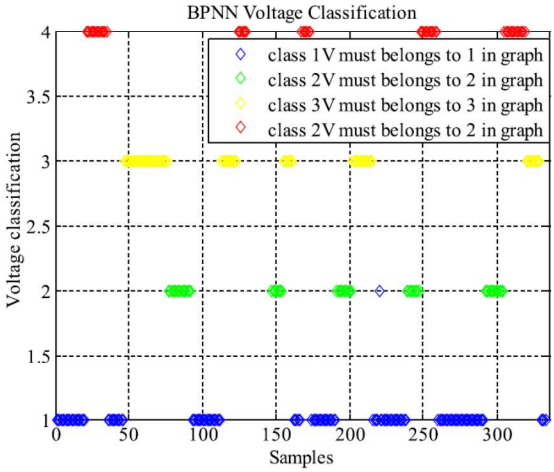


Fig. 7. Vmpp classification using BPNN.

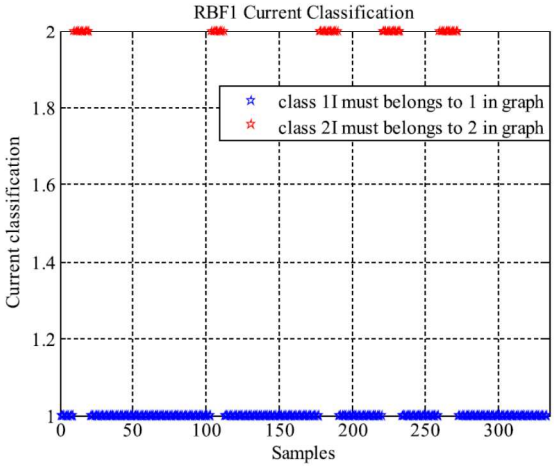


Fig. 9. Impp classification using RBF1.

equal categories which are: 25% for one PV module short circuited, 25% for two PV module short circuited and 25% for four PV module short circuited. For each case there are 700 samples for each attribute. On the other side, the ANN diagnosis employs 336 samples of each attribute (Vmpp (V), Temperature (°C)), which means ($336 \times 2 = 672$ data) distributed in thirty-one cases respectively resume in Table 4.

The obtained classification results of the five algorithms are as follows:

4.1. Detailed analysis of the performance of the five ANNs

4.1.1. BPNN classification

All samples for the classification of the Impp are in their right classes as illustrated by Fig. 6. While nearly all samples for the classification of Vmpp are in their correct classes. Only one confusion case has been encountered during which, the ANN confused a healthy functioning case with a short-circuited PV module as shown by Fig. 7. This confusion is

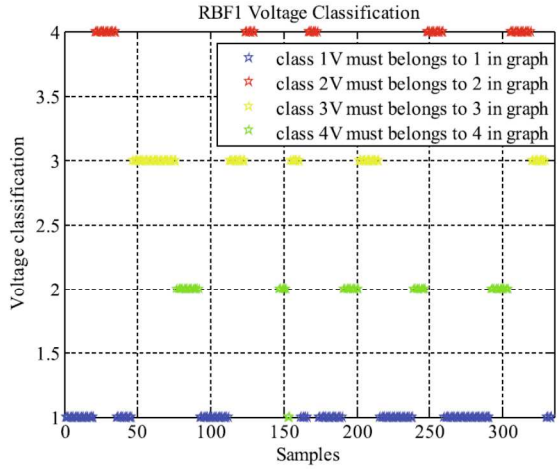


Fig. 10. Vmpp classification using RBF1.

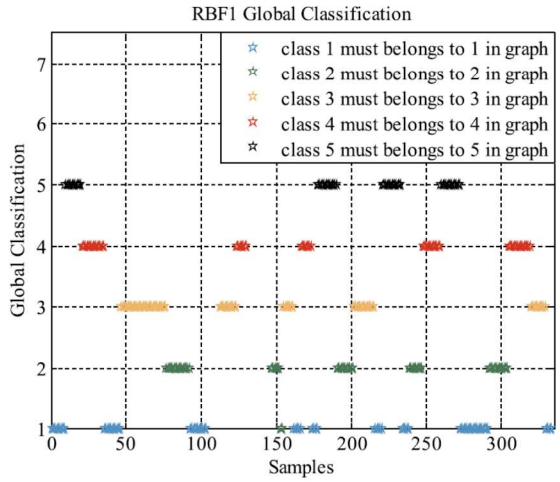


Fig. 11. Global diagnosis of the system using RBF1.

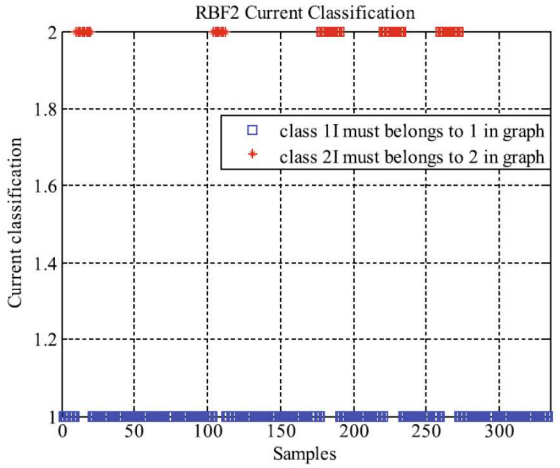


Fig. 12. Impp classification using RBF2.

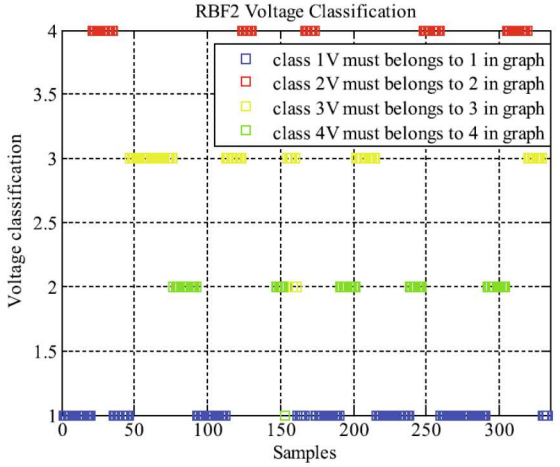


Fig. 13. Vmpp classification using RBF2.

due to the temperature variation. As a result, the combination of the two ANNs classification results outputs reveals a very high accuracy. In fact, only one sample is confused (confusion between C1 and C2) among 336 tested cases as illustrated by Fig. 8.

4.1.2. RBF classification

4.1.2.1. RBF1. The classification results for this type of neural network are demonstrated in Figs. 9 and 10, representing the current classification and the voltage classification respectively. Impp has been successfully classified by RBF1 in any cases while only one confusion case for Vmpp classification has been pointed. The analysis of this confusion case reveals that it is due to the variation of the temperature. As a result the logical combination of the outputs of the two RBF ANNs managed to rank all cases in their right classes except the confusing cases between (C1) and (C2) among the 336 tested cases (see Fig. 11).

4.1.2.2. RBF2. The classification results for the second type of the RBF ANNs are illustrated by Figs. 12 and 13 representing the current and the voltage classification respectively. 100% of Immp are classified in their right classes while Vmpp classification presents three confusion cases. In fact the ANNs confuses between a healty functioning case and one short

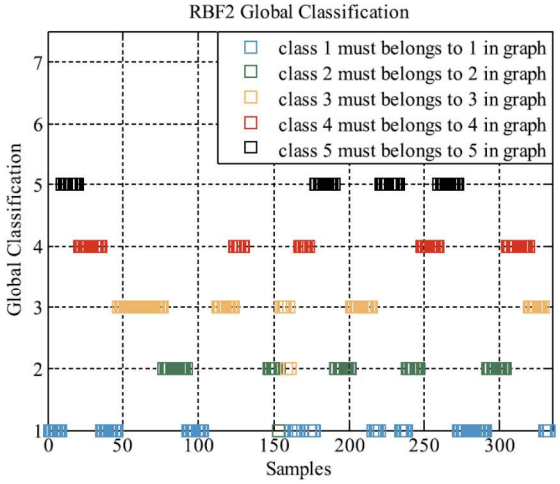


Fig. 14. Global diagnosis of the system using RBF2.

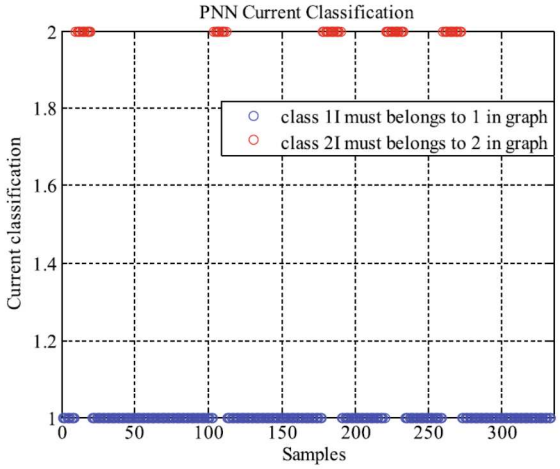


Fig. 15. Impp classification using PNN.

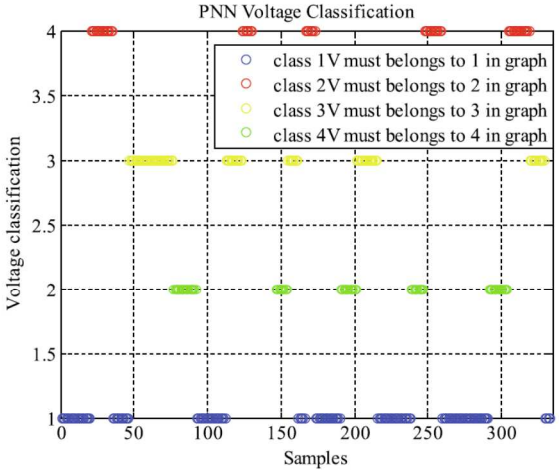


Fig. 16. Vmpp classification using PNN.

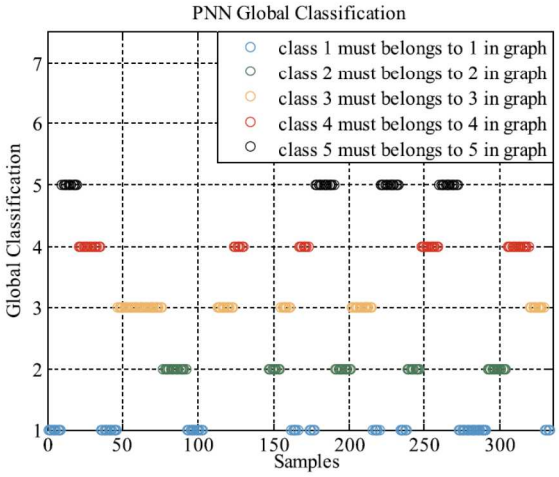


Fig. 17. Global diagnosis of the system using PNN.

circuited PV module and also between two short circuited PV modules and one short circuit PV module. As in the previous confusion cases, the variation of the temperature is the origin of the problem. As consequence, the combination of the ANNs outputs allowed to obtain the results illustrated by Fig. 14. The whole IFD algorithm shows a good accuracy; only three samples are in their wrong classes among 336 samples. The confusions are one sample between (C1) and (C2) and two samples between (C2) and (C3).

4.1.3. PNN classification

All samples of the classification whether for the current or for the voltage are in their right classes without any confusion as illustrated by Figs. 15 and 16 respectively. Fig. 17 shows the diagnosis results of the whole IFD algorithm. The outcome reveals an excellent efficiency.

4.1.4. GRNN classification

The current and voltage classification for the last neural network is demonstrated in Figs. 18 and 19 respectively. All samples for the classification of Impp by GRNN are in their right classes and most samples for the classification of Vmpp by GRNN are in their right classes only two confusion samples are identified. The IFD diagnosis' results are illustrated by Fig. 20 where the outcome reveals a high efficiency, with two confusing cases between (C2) and (C3).

4.2. Comparison of the performance of the five ANNs

For real time diagnosis solution, timeliness of the algorithm has to be taken in consideration. Table 5 lists the running time corresponding to the classification of both electrical parameters and the global diagnosis of PV systems for the five IFD algorithms. It is clear that RBF2 based algorithm can only be used if the diagnosis frequency is greater than one test per 28 min and 20 s while the other algorithms need less one minute to give the diagnosis result.

According to the obtained results, all IFD algorithms present a good accuracy ranging from 98.7% for IFD3 to 100% for IFD4 illustrated in Fig. 21. This little variation is induced by the number of confusion cases which varies from an IFD to another. It is clear that the PNN based algorithm is the best considering this criteria. Regarding the sensitivity, it varies from 96.92% for IFD5 to 100% for IFD4 as shown by Fig. 22. It is clear that this variation is function of the number of the confusion cases

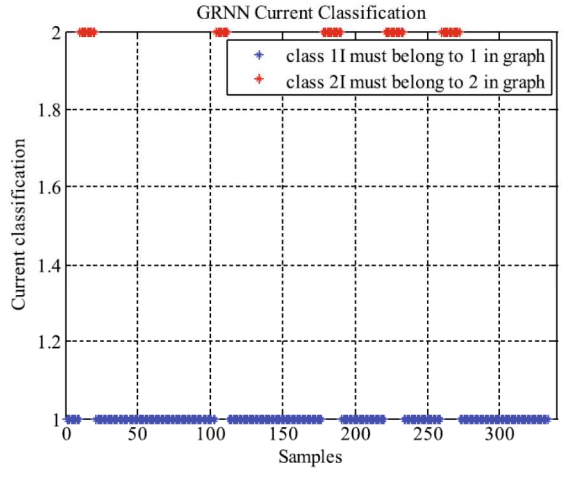


Fig. 18. Impp classification using GRNN.

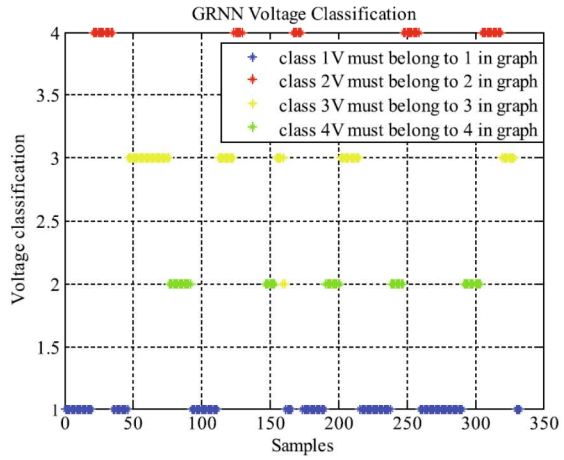


Fig. 19. Vmpp classification using GRNN.

for each classes. This puts the PNN algorithm at the top of the list with zero confusion case and besides thanks to that this algorithm obtains the best results for the precision and sensitivity tests as illustrated by Figs. 23 and 24 respectively.

Fig. 25 shows that globally for accuracy criterion there are not remarkable classes confusion for the five algorithms, while a significant confusion is noticed in class 2 followed by class 1 for sensitivity criterion in all algorithms except IFD4, which does not represent any confusion. For precision and specificity criterions, the confusions are in class 2 and class 3. Additionally, for the four criterions there are no confusion in class 4 and class 5 for the five IFD algorithms. All confusion cases, listed in Table 6 are due to the variation of the temperature.

5. Conclusion

The present paper examines the impact of the Artificial Neural Network choice on the performance of an IFD diagnosis algorithm, which is designed to detect and isolate faults in grid connected PV installation. Four most pertinent criterions are considered in this analysis: the accuracy, the specificity, the sensitivity and the rapidity. The goal is to choose the best ANN to ensure the IFD diagnosis efficiency and

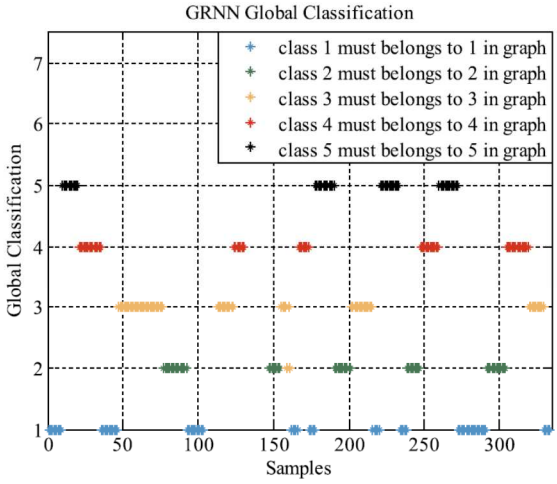


Fig. 20. Global diagnosis of the system using GRNN.

Table 5
The running time for the five IFD algorithms.

	BPNN	RBF1	RBF2	PNN	GRNN
Designation	IFD1	IFD2	IFD3	IFD4	IFD5
Time of (Impp)	13 s	12 s	26min30s	09 s	12 s
Time of (Vmpp)	16 s	28 s	01min36s	15 s	25 s
Time of (Global diagnosis)	34 s	50 s	28min20s	30 s	45 s

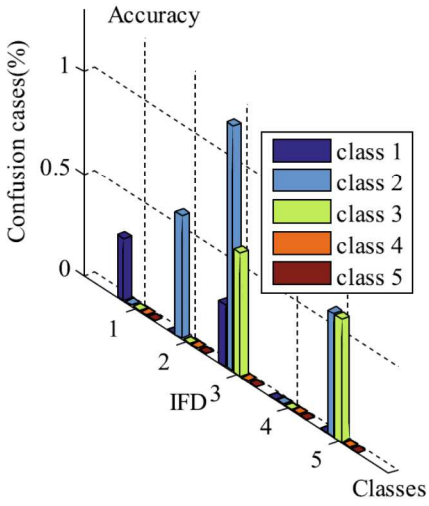


Fig. 21. Confusion of the global accuracy in the five different algorithms.

consequently PV system's safety, durability and reliability. For this propose, five different ANNs have been used and tested: Back Propagation Neural Network, Probabilistic Neural Network, Generalized Regression Neural Network and two Radial Basis Function Neural Network. The performances of these ANNs have been analyzed and compared. The five Neural Network have the same inputs: cell temperature, solar irradiation, voltage and current of the maximum power point of the I-V characteristics.

The IFD algorithm is based on two ANNs to identify and locate the

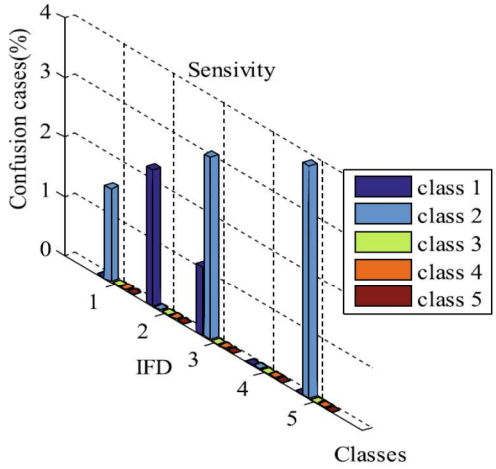


Fig. 22. Confusion of the global sensitivity in the five different algorithms.

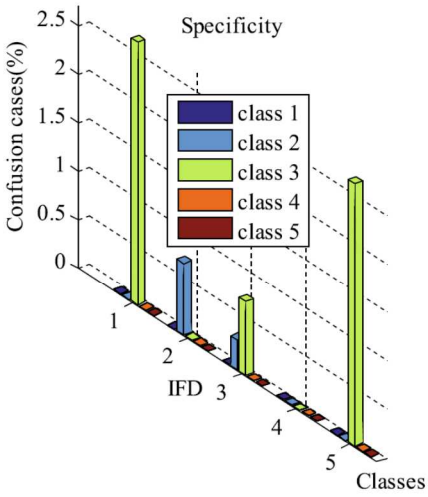


Fig. 24. Confusion of the global specificity in the five different algorithms.

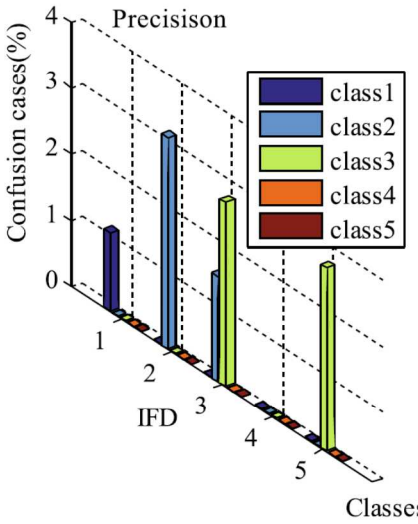


Fig. 23. Confusion of the global precision in the five different algorithms.

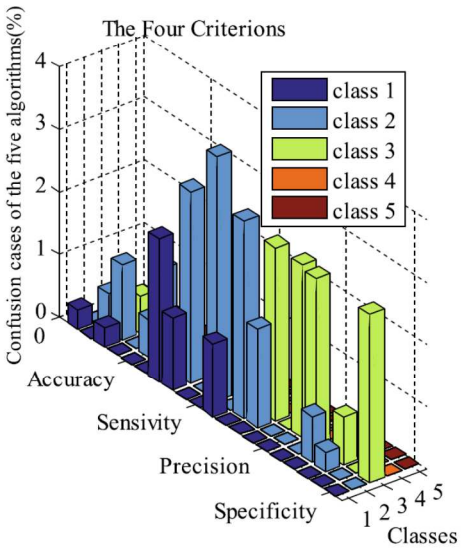


Fig. 25. Confusion of the global criterions in the five different algorithms.

most frequent fault encountered in PV installations: one PV module short circuit in PV string, two PV modules short circuit in PV string, four PV modules short circuit in PV string and one string modules disconnection in PV generator. For their comparison, the five ANNs based IFD algorithms has been tested using 336 different functioning cases enclosing healthy functioning cases and different faulty functioning cases. These investigations demonstrate that the RBF2 based algorithm presents the lowest efficiency with a response time equal to 28 min. noting that its performance varies from 96.82% to 100% on all criterions with three confusion cases for faults classification. While the obtained results from GRNN and RBF1 have presented good results on all criterions with a short response time and a good classification with two confusion classifications cases. Concerning the BPNN based IFD

Table 6
Localization of confusions in all algorithms.

	CONFUSIONS
BPNN	One sample of healthy system in one PV module short circuit.
RBF1	Two sample of one PV module short circuit in healthy system.
RBF2	- One sample of One PV module short circuit in healthy system. - Two samples of two PV modules in one PV module short circuit.
PNN	No confusion.
GRNN	Two samples of two PV modules short circuit in one PV module short circuit.

algorithm, the tests reveal very good results on all criterions varying from 97.27% to 100% with a very good classification's score (one confusion sample). Finally, the PNN based IFD algorithm displaying 100% of success score on all key statistical concepts citing accuracy, sensitivity, specificity and precision compared to other ANNs with no confusion cases. This places it at the top of the list and qualifies it as the best intelligent diagnosis algorithm for the studied grid connected PV installation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Birol F. Solar PV, on track. International Energy Agency; 2020.
[2] Silvestre S, Chouder A, Karatepe E. Automatic fault detection in grid connected PV systems. Sol. Energy 2013;94:119–27.
[3] Chine W, Mellit A, Pavan AM, Kalogirou SA. Fault detection method for grid-connected photovoltaic plants. Renewable Energy 2014;66:99–110.
[4] Platon R, Martel J, Woodruff N, Chau TY. Online fault detection in PV systems. IEEE Trans Sustainable Energy 2015;6(4):1200–7.
[5] Drews A, de Keizer AC, Beyer HG, Lorenz E, Betcke J, van Sark WGJHM, et al. Monitoring and remote failure detection of grid connected PV systems based on satellite observations. Sol Energy 2007;81:548–64.
[6] Dhoke A, Sharma R, Kumar ST. An approach for fault detection and location in solar PV systems. Sol Energy 2019;194:197–208.
[7] Garoudja E, Harrou F, Sun Y, Kara K, Chouder A, Silvestre S. Statistical fault detection in photovoltaic systems. Sol Energy 2017;150:485–99.
[8] Chen Z, Chen Y, Wu L, Cheng S, Lin P. Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. Energy Convers Manage 2019;198:111793.
[9] Solórzano J, Egido MA. Automatic fault diagnosis in PV systems with distributed MPPT. Energy Convers Manage 2013;76:925–34.
[10] Tina G.P, Cosentino F, Ventura F. Monitoring and Diagnostics of Photovoltaic Power Plants. World Renewable Energy Congress; 2014, London, United Kingdom.
[11] Spataru S, Sera D, Kerekes T, Teodorescu R. Diagnostic method for photovoltaic systems based on light I-V measurements. Sol Energy 2015;119:29–44.
[12] Hare J, Shi X, Gupta S, Bazzi A. Fault diagnostics in smart micro-grids: a survey. Renewable Sustainable Energy Rev 2016;60:1114–24.
[13] Silvestre S, daSilva M, Chouder A, Guasch D, Karatepe E. New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators. Energy Convers Manage 2014;86(2):41–9.
[14] Tadj M, Benmouiza K, Cheknane A, Silvestre S. Improving the performance of PV systems by faults detection using GISTEL approach. Energy Convers Manage 2014; 80:298–304.
[15] Madeti SR, Singh SN. Modeling of PV system based on experimental data for fault detection using kNN method. Sol Energy 2018;173:139–51.
[16] Hussain M, Dhimish M, Titarenko S, Mather P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. Renewable Energy 2020;155:1272–92.
[17] Belaout A, Krim F, Mellit A, Talbi B, Arabi A. Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification. Renewable Energy 2018;127:548–58.
[18] Liao Z, Wang D, Tang L, Ren J, Liu Z. A Heuristic diagnostic method for a PV system: triple-layered particle swarm optimization–back-propagation neural network. Energies 2017;10(2):226. https://doi.org/10.3390/en10020226.
[19] Garoudja E, Chouder A, Kara K, Silvestre S. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. En Con Man 2017;151: 496–513.
[20] Zhu H, Lub L, Yao J, Daia S, Hu Y. Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. Sol Energy 2018;176:395–405.
[21] Kara Mostefa Khelil C, Amrouche B, Benyoucef AS, Kara K, Chouder A. New Intelligent Fault Diagnosis (IFD) Approach for grid-connected photovoltaic systems. J Energy 2020;211. 118591.
[22] Harrou F, Fillatre L, Nikiforov I. Anomaly detection/detectability for a linear model with a bounded nuisance parameter. Ann Rev Contr 2014;38(1):32–4.
[23] Chouder A, Silvestre S. Automatic supervision and fault detection of PV systems based on power losses analysis. Energy Convers Manag 2010;51:1929–37.
[24] Gokmen N, Karatepe E, Celik B, Silvestre.. Simple diagnostic approach for determining of faulted PV modules in string based PV arrays. Sol Energy 2012;86: 3364–77.
[25] Chouder A, Silvestre S. Analysis model of mismatch power losses in PV systems. Sol Energy 2009;131(2):024504–24514.
[26] Bastidas-Rodriguez JD, Franco E, Petrone G, Ramos-Paja CA, Spagnuolo G. Quantification of photovoltaic module degradation using model based indicators. Math Comput Simul 2017;131:101–13.
[27] Takashima T, Yamaguchi J, Ishida M. Fault detection by signal response in PV module strings. In: Proceedings of the 33rd IEEE Photovoltaic Specialists Conference; 11–16 May 2008: 1–5.
[28] Johnson J, Kuszmaul S, Bower W, Schoenwald D. Using PV module and line frequency response data to create robust arc fault detectors. In: Proceedings of the 26th European Photovoltaic Solar Energy Conference and Exhibition; 05–09 September 2011, Hamburg, Germany; 3745–50.
[29] Takashima T, Yamaguchi J, Otani K, Oozeki T, Kato K, Ishida M. Experimental studies of fault location in PV module strings. Sol Energy Mater Sol Cel 2009;93(6-7):1079–82.
[30] Hachana O, Tina GM, Hemsas KE, Tina Giuseppe Marco, Hemsas K. PV array fault diagnostic technique for BIPV Systems. Energy Build 2016;126:263–74.
[31] Zhao Y, dePalma J, Mosesian J, Lyons R, Lehman B. Line–line fault analysis and protection challenges in solar photovoltaic arrays. IEEE Trans. Ind. Electron 2013; 60(9):3784–95.
[32] SolarEdge. Performance of PV Topologies under Shaded Conditions (white paper). SolarEdge; 2013.
[33] Kjær S.B, Oprea O, Borup U. Adaptive sweep for PV applications. In: 26th European Photovoltaic Solar Energy Conference and Exhibition. Hamburg, Germany; 2011: 3708–10.
[34] Swingler A. Photovoltaic String Inverters and Shade-Tolerant Maximum Power Point Tracking: Toward Optimal Harvest Efficiency and Maximum ROI (white paper). Burnaby, Canada: Schneider Electric; 2010.
[35] Suganthi L, Iniyan S, Samuel A. Applications of fuzzy logic in renewable energy systems – a review. Renewable Sustainable Energy Rev 2015;48:585–607.
[36] Ducange P, Fazzolari M, Lazzerini B, Marcelloni F. An intelligent system for detecting faults in photovoltaic fields. In: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA); 22–24 November 2011, Cordoba: 1341–46.
[37] Bonsignore L, Davarifar M, Rabhi A, Tina GM, Elhajjaji A. Neuro-Fuzzy fault detection method for photovoltaic systems. Energy Procedia 2014;62:431–41.
[38] Dhimish M, Holmes V, Mehrdadi B, Dales M. Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection. Renewable Energy 2018;117: 257–74.
[39] Zhao Y, Yang L, Lehman B, De Palma JF, Mosesian J, Lyons R. Decision based fault detection and classification in solar photovoltaic arrays. In: Twenty Seventh Annual IEEE Applied Power Electronics Conference and Exposition; 2012. p. 93–9.
[40] Mekki H, Mellit A, Salhi H. Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules. Simul Model Pract Theory 2016; 67:1–13.
[41] Yuchuan W, Qinli L, Yaqin S. Application of BP neural network fault diagnosis in solar Photovoltaic System, in: Proceedings of the IEEE International Conference on Mechatronics and Automation, Changchun, China; 2009: 9–12.
[42] Syafaruddin S, Karatepe E, Hiyama T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In: Proceedings of the 16th International Conference on Intelligent System Application to Power Systems (ISAP), Greece; 2011:1-6.
[43] Li Z, Wang Y, Zhou D, Wu C. An intelligent method for fault diagnosis in photovoltaic array, ICSC Part II CCIS 2012;327:10-6.
[44] Karatepe E, Hiyama T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In: 2011 16th international conference on intelligent system application to power systems (ISAP). IEEE; 2011. p. 1–6.
[45] Chine W, Mellit A, Lughi V, Malek A, Sulligoi G, Massi Pavan A. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renewable Energy 2016;90:501–12.
[46] Ahmadipour M, Hizam H, Othman ML, Radzi MAM, Murthy AS. Avinash Srikantaislanding detection technique using Slantlet Transform and Ridgelet Probabilistic Neural Network in grid-connected photovoltaic system. J Appl Energy 2018;231:645–59.
[47] Amrouche B, Pivert X. Artificial neural network based daily local forecasting for global solar radiation. J Appl Energy 2014;130:333–41.
[48] Liang L, Guo W, Zhang Y, Zhang W, Li L, Xing X. Radial Basis Function Neural Network for prediction of medium frequency sound absorption coefficient of composite structure open-cell aluminum foam. Appl Acoust 2020;170:107505. https://doi.org/10.1016/j.apacoust.2020.107505.
[49] Matera F. Radial basis function neural network. Subst Use Misuse 1998;33(2): 317–34.
[50] Hong H, Zhang Z, Guo A, Shen L, Sun H, Liang Y, et al. Radial basis function artificial neural network (RBF ANN) as well as the hybrid method of RBF ANN and

grey relational analysis able to well predict trihalomethanes levels in tap water. J Hydrol 2020;591:125574.
[51] <https://fr.mathworks.com/help/deeplearning/ref/newrb.html>.
[52] Specht DF. Probabilistic neural networks for classification, mapping, or associative memory. In: IEEE international conference on neural networks; 1988. p. 525–32.
[53] Fan H, Pei J, Zhao Y. An optimized probabilistic neural network with unit hyperspherical crown mapping and adaptive kernel coverage. Neurocomputing 2020;373:24–34.
[54] Specht DF. A general regression neural network. IEEE Trans Neural Networks 1991;2(6):568–76.
[55] Yuan Q, Xu H, Li T, Shen H, Zhang L. Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental U.S. J Hydrology 2020;580:124351.
[56] <https://en.wikipedia.org/wiki/Sensitivity_and_specificity>.

Chérifa Kara Mostefa Khelil[a,c,*], Badia Amrouche[b,c], Kamel Kara[c], Aissa Chouder[d]
[a] *Electrical Engineering Department, Khemis Miliana University, Ain Defla, Algeria*
[b] *Renewable Energies Department, Blida 1 University, BP 270 Blida, Algeria*
[c] *SET Laboratory, Electronics Department, Blida 1 University, BP 270 Blida, Algeria*
[d] *Electrical Engineering Laboratory (LGE), University Mohamed Boudiaf of M'sila, BP 166, 28000, Algeria*

[*] Corresponding author.
E-mail address: k.karamostapha@univ-dbkm.dz (C. Kara Mostefa Khelil).